CORPORATE INSTITUTE OF MANAGEMENT, BHOPAL

Subject Name: (QUANTITATIVE METHODS) Subject code: CP-102

Lecture Notes By - Prof. Akhilesh jain

Course : MBA , Semester: I

"Education to All for Exc

Unit - I	Statistical basis of managerial decision: Frequency distribution and graphic representation of frequency distribution, Measures of Central Tendency - Mean, Median, Mode, Requisite of ideal measures of Central techniques, Merits, Domestic of Mean, Median Mode and their managerial application.

1.1 INTRODUCTION

Data that have not been organised in any way are called raw data. They are collected by counting or measurement, or through other survey methods.

The next step after the completion of data collection is to organize the data into a meaningful form so that a trend, if any, emerging out of the data can be seen easily. One of the common methods for organizing data is to construct frequency distribution. Frequency distribution is an organized tabulation/graphical representation of the number of individuals in each category on the scale of measurement.[1] It allows the researcher to have a glance at the entire data conveniently. It shows whether the observations are high or low and also whether they are concentrated in one area or spread out across the entire scale. Thus, frequency distribution presents a picture of how the individual observations are distributed in the measurement scale.

A *frequency distribution* is an organized tabulation of the number of individuals located in each category on the scale of measurement.

A frequency distribution can be structured either as a table or as a graph, but in either case the distribution presents the same two elements:

1. The set of categories that make up the original measurement scale.

2. A record of the frequency, or number of individuals, in each category.

1.2 DISPLAYING FREQUENCY DISTRIBUTIONS

Frequency tables:

A frequency (distribution) table shows the different measurement categories and the number of observations in each category. Before constructing a frequency table, one should have an idea about the range (minimum and maximum values). The range is divided into arbitrary intervals called “class interval.” If the class intervals are too many, then there will be no reduction in the bulkiness of data and minor deviations also become noticeable. On the other hand, if they are very few, then the shape of the distribution itself cannot be determined.

The width of the class can be determined by dividing the range of observations by the number of classes. The following are some guidelines regarding class widths:

- It is advisable to have equal class widths. Unequal class widths should be used only when large gaps exist in data.

- The class intervals should be mutually exclusive and non overlapping.
- Open-ended classes at the lower and upper side (e.g., <10, >100) should be avoided.

The frequency distribution table of the resting pulse rate in healthy individuals is given in Table 1. It also gives the cumulative and relative frequency that helps to interpret the data more easily.

Table 1 : Frequency distribution of the resting pulse rate in healthy volunteers (N = 63)

Pulse/min	**Frequency**	**Cumulative frequency**	**Relative cumulative frequency (%)**
60–64	2	2	3.17
65–69	7	9	14.29
70–74	11	20	31.75
75–79	15	35	55.56
80–84	10	45	71.43
85–89	9	54	85.71
90–94	6	60	95.24
95–99	3	63	100

1.3 GRAPHICAL REPRESENTATION OF A FREQUENCY DISTRIBUTION:

A graph is a method of presenting statistical data in visual form. The main purpose of any chart is to give a quick, easy-to-read-and-interpret pictorial representation of data which is more difficult to obtain from a table or a complete listing of the data. The type of chart or graphical presentation used and the format of its construction is incidental to its main purpose. A well-designed graphical presentation can effectively communicate the data’s message in a language readily understood by almost everyone. You will see that graphical methods for describing data are intuitively appealing descriptive techniques and that they can be used to describe either a sample or a population; quantitative or qualitative data sets.

Some basic rules for the construction of a statistical chart are listed below:

(a) Every graph must have a clear and concise title which gives enough identification of the graph.

(b) Each scale must have a scale caption indicating the units used.

(c) The zero point should be indicated on the co-ordinate scale. If, however, lack of space makes it inconvenient to use the zero point line, a scale break may be inserted to indicate its omission.

(d) Each item presented in the graph must be clearly labelled and legible even in black and white reprint.

There are many varieties of graphs. The most commonly used graphs are described as below.

1. Line diagram
2. Histogram
3. Bar diagram

4. Pie chart
5. Frequency polygon
6. Ogives or Cumulative frequency graphs

Let us discuss the above graphical representations of frequency distribution in detail.

1.3.1 LINE DIAGRAM

When the time series exhibit a wide range of fluctuations, we may think of logarithmic or ratio chart where "Log y" and not "y" is plotted against "t".

We use Multiple line chart for representing two or more related time series data expressed in the same unit and multiple – axis chart in somewhat similar situations, if the variables are expressed in different units.

Example : *The profits in thousands of dollars of an industrial house for 2002, 2003, 2004, 2005, 2006, 2007 and 2008 are 5, 8, 9, 6, 12, 15 and 24 respectively. Represent these data using a suitable diagram.*

Solution : We can represent the profits for 7 consecutive years by drawing either a line diagram as given below.

Let us consider years on horizontal axis and profits on vertical axis.

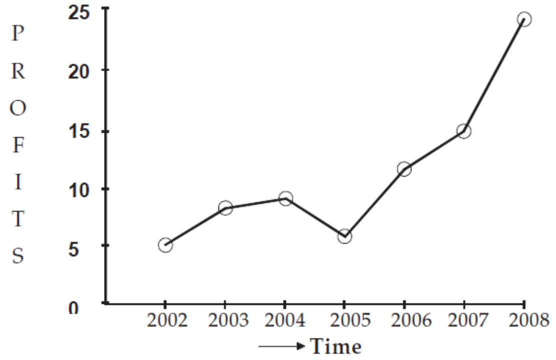
For the year 2002, the profit is 5 thousand dollars. It can be written as a point (2002, 5)

In the same manner, we can write the following points for the succeeding years.

(2003, 8), (2004, 9), (2005, 6), (2006, 12), (2007, 15) and (2008, 24)

Now, plotting all these point and joining them using ruler, we can get the line diagram.

Showing line diagram for the profit of an Industrial House during 2002 to 2008.



Let us look at the next stuff on "Graphical representation of a frequency distribution"

1.3.2 HISTOGRAM

A two dimensional graphical representation of a continuous frequency distribution is called a histogram.

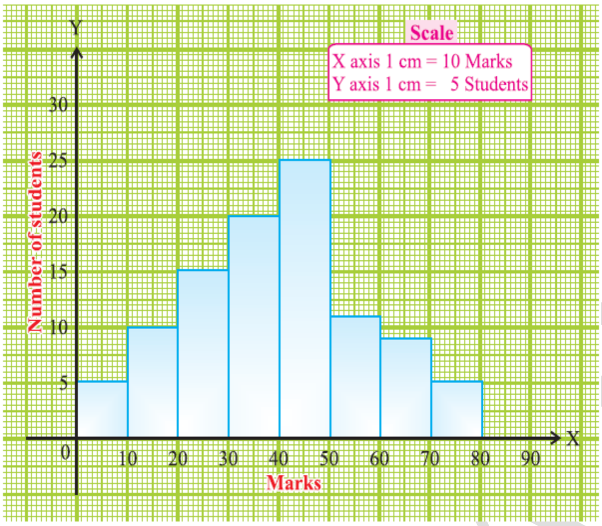
In histogram, the bars are placed continuously side by side with no gap between adjacent bars.

That is, in histogram rectangles are erected on the class intervals of the distribution. The areas of rectangle are proportional to the frequencies.

Example : Draw a histogram for the following table which represent the marks obtained by 100 students in an examination :

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. Of Students	5	10	15	20	25	12	8	5

Solution : The class intervals are all equal with length of 10 marks.
Let us denote these class intervals along the X-axis.
Denote the number of students along the Y-axis, with appropriate scale.
The histogram is given below.



Let us look at the next stuff on "Graphical representation of a frequency distribution".

1.3.3 BAR DIAGRAM

There are two types of bar diagrams namely, Horizontal Bar diagram and Vertical bar diagram.

While horizontal bar diagram is used for qualitative data or data varying over space, the vertical bar diagram is associated with quantitative data or time series data.

Bars i.e. rectangles of equal width and usually of varying lengths are drawn either horizontally or vertically.

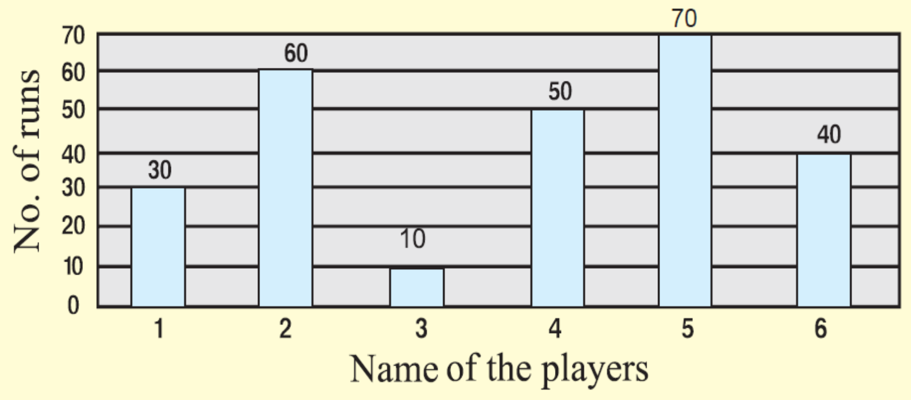
We consider Multiple or Grouped Bar diagrams to compare related series. Component or sub-divided Bar diagrams are applied for representing data divided into a number of components. Finally, we use Divided Bar charts or Percentage

Bar diagrams for comparing different components of a variable and also the relating of the components to the whole. For this situation, we may also use Pie chart or Pie diagram or circle diagram.

Example: The total number of runs scored by a few players in one-day match is given.

Players	1	2	3	4	5	6
No of Runs	30	60	10	50	70	40

Solution : Draw bar graph for the above data.



Let us look at the next stuff on "Graphical representation of a frequency distribution"

1.3.4 PIE CHART

In a pie chart, the various observations or components are represented by the sectors of a circle and the whole circle represents the sum of the value of all the components .Clearly, the total angle of 360° at the center of the circle is divided according to the values of the components.

The central angle of a component is = [Value of the component / Total value] x 360°

Sometimes, the value of the components is expressed in percentages.

In such cases,

The central angle of a component is = [Percentage value of the component / 100] x 360°

Example : The number of hours spent by a school student on various activities on a working day, is given below. Construct a pie chart using the angle measurement.

Activity	Sleep	School	Play	Home Work	Other
No. Of Hours	8	6	3	3	4

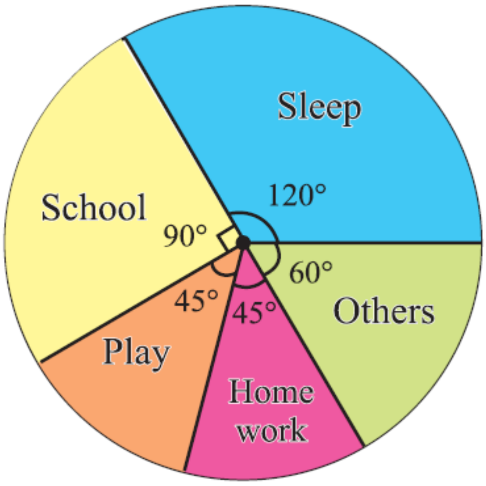
Draw a pie chart to represent the above information.

Solution : The central angle of a component is = [Value of the component / Total value] x 360°

We may calculate the central angles for various components as follows :

Activity	**No. Of Hours**	Central Angle
Sleep	8	$8/24 * 360^0 = 120^0$
School	6	$6/24 * 360^0 = 90^0$
Play	3	$3/24 * 360^0 = 45^0$
Home Work	3	$3/24 * 360^0 = 45^0$
Other	4	$4/24 * 360^0 = 60^0$
Total	24	360^0

From the above table, clearly, we obtain the required pie chart as shown below.



Let us look at the next stuff on "Graphical representation of a frequency distribution"

1.3.5 FREQUENCY POLYGON

Frequency Polygon is another method of representing frequency distribution graphically.

Obtain the frequency distribution and compute the mid points of each class interval.

Represent the mid points along the X-axis and the frequencies along the Y-axis.

Plot the points corresponding to the frequency at each mid point.

Join these points, by straight lines in order.

To complete the polygon join the point at each end immediately to the lower or higher class marks (as the case may be at zero frequency) on the X-axis.

Example: Draw a frequency polygon for the following data without using histogram.

Class Interval	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90
Frequency	4	6	8	10	12	14	7	5

Solution: Mark the class intervals along the X-axis and the frequency along the Y-axis.

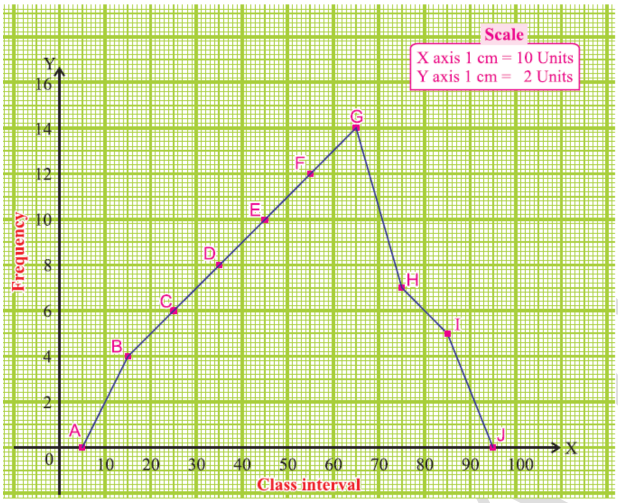
We take the imagined classes 0-10 at the beginning and 90-100 at the end, each with frequency zero.

We have tabulated which is given below.

Class Interval	Mid Point	Frequency
0-10	5	0
10-20	15	4
20-30	25	6
30-40	35	8
40-50	45	10
50-60	55	12
60-70	65	14
70-80	75	7
80-90	85	5
90-100	95	0

Using the adjacent table, plot the points A (5, 0), B (15, 4), C (25, 6), D (35, 8), E (45, 10), F (55, 12), G (65, 14), H (75, 7), I (85, 5) and J (95, 0).

We draw the line segments AB, BC, CD, DE, EF, FG, GH, HI, IJ to obtain the required frequency polygon ABCDEFGHIJ, which is given below.



Let us look at the next stuff on "Graphical representation of a frequency distribution"

1.4 FREQUENCY CURVE: A frequency curve can be obtained by smoothing the frequency polygon.

1.4.1 Ogives or Cumulative frequency graphs

By plotting cumulative frequency against the respective class boundary, we get ogives.

As such there are two ogives – less than type ogives, obtained by taking less than cumulative frequency on the vertical axis and more than type ogives by plotting more than type cumulative frequency on the vertical axis and thereafter joining the plotted points successively by line segments.

Example : Draw ogives for the following table which represents the frequency distribution of weights of 36 students.

Weight in Kg.	43.50-48.50	48.50-53.50	53.50-58.50	58.50-63.50	63.50-68.50	68.50-73.50
No. of Students	3	4	5	7	9	8

Solution : To draw ogives for the above frequency distribution, we have to write less than and more than cumulative frequency as given below.

Weight in Kg.	**No. of Students**	**Weight in Kg (CB)**	**Commulative Frequency**	
			less than	**More than**
43.50-48.50	3	43.5	3	0+8+9+7+5+4+3=36
48.50-53.50	4	48.5	0+3=3	0+8+9+7+5+4=33
53.50-58.50	5	53.5	0+3+4=7	0+8+9+7+5=29
58.50-63.50	7	58.5	0+3+4+5=12	0+8+9+7=24
63.50-68.50	9	63.5	0+3+4+5+7=19	0+8+9=17
68.50-73.50	8	68.5	0+3+4+5+7+9=28	0+8=8
		730.5	0+3+4+5+7+9+8=36	0

Now, we have to write the points from less than and more than cumulative frequency as given below.

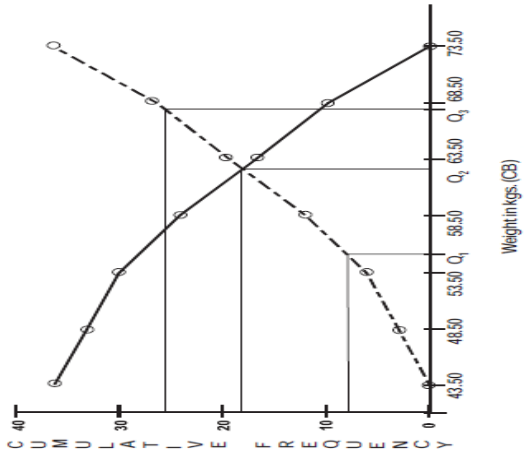
Points from less than cumulative frequency :

(43.50, 0), (48.50, 3), (53.50, 7), (58.50, 12), (63.50, 19), (68.50, 28) and (73.50, 36)

Points from more cumulative frequency :

(43.50, 36 (48.50, 33), (53.50, 29), (58.50, 24), (63.50, 17), (68.50, 8) and (73.50, 0)

Now, taking frequency on the horizontal axis, weights on vertical axis and plotting the above points, we get ogives as given below.



After having gone through the stuff given above, we hope that the students would have understood "Graphical representation of a frequency distribution".

1.5 RELATIVE FREQUENCY

Relative frequency of a class is defined as: $\dfrac{\text{Frequency of the Class}}{\text{Total Frequency}}$

If the frequencies are changed to relative frequencies, then a relative frequency histogram, a relative frequency polygon and a relative frequency curve can similarly be constructed.

Relative frequency curve can be considered as probability curve if the total area under the curve be set to 1. Hence the area under the relative frequency curve between a and b is the probability between interval a and b.

1.6 CENTRAL TENDENCY:

When we work with numerical data, it seems apparent that in most set of data there is a tendency for the observed values to group themselves about some interior values; some central values seem to be the characteristics of the data. This phenomenon is referred to as central tendency. For a given set of data, the measure of location we use depends on what we mean by middle; different definitions give rise to different measures. We shall consider some more commonly used measures, namely arithmetic mean, median and mode. The formulas in finding these values depend on whether they are ungrouped data or grouped data.

1.7 Measures of Central Tendency

Measures of central tendency, or more simply measures of centre, indicate where the centre or most typical value of a data set lies.

Average:

The term average is used to express an amount that is typical for a group of people or things. For example, you may read in a newspaper that on average people watch 3 hours of television per day. We understand from the use of the term average that not everybody watches 3 hours of television each day, but that some watch more and some less. However, we realize from the use of the term average that the figure of 3 hours per day is a good indicator of the amount of TV watched in general.

Averages are useful because they:

- Summaries a large amount of data into a single value
- Indicate that there is some variability around this single value within the original data.

In everyday language most people have an inherent understanding of what the term average means. However, within the language of mathematics there are three different definitions of average known as the mean, median and mode.

TYPES OF AVERAGE: Average or measure of central tendency is of following types:

1. Mathematical Average
    - (i) Arithmetic mean
    - (ii) Geometric mean
    - (iii) Harmonic mean
2. Potential Average
    - (i) Median
    - (ii) Mode

The Term "Average" is incorrect the correct term should be "Central Tendency".

1.7.1 Arithmetic Mean

Arithmetic Mean is defined as the sum of observations divided by the number of observations. It can be computed in two ways:

1) Simple arithmetic mean and
2) Weighted arithmetic mean.

In case of simple arithmetic mean, equal importance is given to all the observations while in weighted arithmetic mean, the importance given to various observations is not same.

Calculation of Simple Arithmetic Mean

(a) When Individual Observations are given.

Let there be n observations $X_1, X_2..... X_n$. Their arithmetic mean can be calculated either by direct method or by short cut method. The arithmetic mean of these observations will be denoted by X

1.7.1.1 Direct Method: Under this method, X is obtained by dividing sum of observations by number of observations, i.e.,

$$\text{Mean of a set of sample values} \quad \bar{X} = \frac{\sum x}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

[Where Σ - denotes *summation* of a set of values , x – is the *variable* usually used to represent the individual data values, n – represents the *number of values in a sample]*

(2) Mean from a frequency table $\mu = \frac{\sum f_i \cdot x_i}{\sum f_i}$

Where - for **grouped data**: f_i - is the frequency in the i^{th} class, x_i - is the class mark in the i^{th} class;

for **ungrouped data**: f_i - is the frequency in the i^{th} datum x_i - is the value in the i^{th} datum.

Arithmetic mean can be used to calculate any numerical data and it is always unique. It is obvious that extreme values affect the mean. Also, arithmetic mean ignores the degree of importance in different categories of data.

Example : Given the following set of ungrouped data: 20, 18, 15, 15, 14, 12, 11, 9, 7, 6, 4, 1

Find the mean of the ungrouped data.

Solution : $mean = \frac{20+18+2\cdot 15+14+12+11+9+7+6+4+1}{12} = \frac{132}{12} = 11$

1.7.1.2 Short-cut Method: This method is used when the magnitude of individual observations is large. The use of short-cut method is helpful in the simplification of calculation work. Let A be any assumed mean. We subtract A from every observation. The difference between an observation and A, i.e., $Xi - A$ is called the deviation of i th observation from A and is denoted by di. Thus, we can write ;

$$\bar{X} = A + \frac{\sum d_i}{n} = A + \frac{\sum (X_i - A)}{n}$$

This result can be used for the calculation of X .

Remarks: Theoretically we can select any value as assumed mean. However, for the purpose of simplification of calculation work, the selected value should be as nearer to the value of X as possible.

Example : The following figures relate to monthly output of cloth of a factory in a given year

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Out put (ooo meters)	80	88	92	84	96	92	96	100	92	94	98	86

Calculate the average monthly output.

Solution:

(i) Using Direct Method:

$$\bar{X} = \frac{80+88+92+84+96+92+96+100+92+94+98+86}{12} = 91.5(000\,mtrs)$$

(ii) **Using Shortcut Method:** Let Assumed mean is A=90

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	
Out put (ooo meters)	80	88	92	84	96	92	96	100	92	94	98	86	Total
$d_i=Xi-A$	-10	-2	2	-6	6	2	6	10	2	4	8	-4	$\Sigma d_i=18$

$$\bar{X} = 90 + \frac{18}{2} = 91.5(\text{thousand}\,mtrs)$$

(b) When data are in the form of an ungrouped frequency distribution

Let there be n values X_1 , X_2,.................X_n out of which X_1 has occured f_1 times , X_2 has occured f_2 times, ... X_n has occured f_n times . Let N be the total frequency, i.e. $N=\Sigma f_i$

Alternatevly

Values	X_1	X_2	....	X_n
Frequency	f_1	f_2	./..............	f_n

Direct Method : $$\bar{X} = \frac{X_1.f_1 + X_2.f_2 + ...+ X_n.f_n}{f_1 + f_2 + ...+ f_n} = \frac{\sum X_i.f_i}{\sum f_i} = \frac{\sum X_i.f_i}{N}$$

Short Cut Method: $$\bar{X} = A + \frac{\sum f_i d_i}{\sum f_i} = \frac{\sum f_i d_i}{N} = A + \bar{d}$$

Example: The following is the frequency distribution of age of 670 students of a school. Compute the arithmetic mean of the data.

X(years)	5	6	7	8	9	10	11	12	13	14
Frequency	25	45	90	165	112	96	81	26	18	12

Solution:

X(years)	5	6	7	8	9	10	11	12	13	14	Total
Frequency (*f*)	25	45	90	165	112	96	81	26	18	12	$\Sigma f=670$
f x	125	270	630	1320	1008	960	891	312	234	168	$\Sigma fx=5918$
d=X-8	-3	-2	-1	0	1	2	3	4	5	6	
fd	-75	-90	-90	0	112	192	243	104	90	72	$\Sigma fd=558$

Direct Method: $$\bar{X} = \frac{X_1.f_1 + X_2.f_2 + ...+ X_n.f_n}{f_1 + f_2 + ...+ f_n} = \frac{\sum X_i.f_i}{\sum f_i} = \frac{\sum X_i.f_i}{N} = \frac{5918}{670} = 8.83\,years$$

Short Cut Method :

$$\bar{X} = A + \frac{\sum f_i d_i}{\sum f_i} = \frac{\sum f_i d_i}{N} = A + \bar{d} = 8 + \frac{558}{670} = 8.83 years$$

(c) When data are in the form of a grouped frequency distribution

In a grouped frequency distribution, there are classes along with their respective frequencies. Let l_i be the lower limit and ui be the upper limit of i^{th} class. Further, let the number of classes be n, so that i = 1, 2,.....n. Also let fi be the frequency of i^{th} class. This distribution can write in tabular form, as shown.

Note: Here u_1 may or may not be equal to l_2, i.e., the upper limit of a class may or may not be equal to the lower limit of its following class.

It may be recalled here that, in a grouped frequency distribution, we only know the number of observations in a particular class interval and not their individual magnitudes. Therefore, to calculate mean, we have to make a fundamental assumption that the observations in a class are uniformly distributed.

Under this assumption, the mid-value of a class will be equal to the mean of observations in that class and hence can be taken as their representative. Therefore, if X_i is the mid-value of i^{th} class with frequency f_i, the above assumption implies that there are f_i observations each with magnitude X_i (i = 1 to n). Thus, the arithmetic mean of a grouped frequency distribution can also be calculated by the use of the formula, given in § below.

Values	l_1-u_1	l_2-u_2	....	l_2-u_2	*Total*
Frequency (*f*)	f_1	f_2	./...............	f_n	$\Sigma f_i = N$

Remarks: The accuracy of arithmetic mean calculated for a grouped frequency distribution depends upon the validity of the fundamental assumption. This assumption is rarely met in practice. Therefore, we can only get an approximate value of the arithmetic mean of a grouped frequency distribution.

Example: Calculate arithmetic mean of the following distribution:

Class Interval	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Frequency	3	8	12	15	18	16	11	5

Solution: Here only short-cut method will be used to calculate arithmetic mean but it can also be calculated by the use of direct-method

Class Interval	**Frequency (*f*)**	**Mid Value (*X*)**	***d*=*X*-35**	***fd***
0-10	3	5	-30	-90
10-20	8	15	-20	-160
20-30	12	25	-10	-120

30-40	15	35	0	0
40-50	18	45	10	180
50-60	16	55	20	320
60-70	11	65	30	330
70-80	5	75	40	200
Total	88			660

$$\bar{X} = A + \frac{\sum fd}{n} = 35 + \frac{660}{88} = 42.5$$

Example : The following table gives the distribution of weekly wages of workers in a factory. Calculate the arithmetic mean of the distribution.

Weekly Wages	240-269	270-299	300-329	330-359	360-389	390-419	420-449
No. Of workers	7	19	27	15	12	12	8

Solution*:* It may be noted here that the given class intervals are inclusive. However, for the computation of mean, they need not be converted into exclusive class intervals.

class intervals	**Mid values (*X*)**	**Frequency (*f*)**	***d=X-A=X-344.5***	***fd***
240-269	254.5	7	-90	-630
270-299	284.5	19	-60	-1140
300-329	314.5	27	-30	-810
330-359	344.5	15	0	0
360-389	374.5	12	30	360
390-419	404.5	12	60	720
420-449	434.5	8	90	720
	Total	100		-780

$$\bar{X} = A + \frac{\sum fd}{N} = 344.5 - \frac{780}{100} = 336.7$$

(c) STEP DEVIATION METHOD OR CODING METHOD

In a grouped frequency distribution, if all the classes are of equal width, say *'h'*, the successive mid-values of various classes will differ from each other by this width. This fact can be utilised for reducing the work of computations.

Let us define $u_i = \frac{X_i - A}{h}$ then $\bar{X} = A + h\frac{\sum_{i=1}^{n} f_i u_i}{N}$

Using this relation we can solve above problem such that

class intervals	**Mid values (*X*)**	**Frequency (*f*)**	***d=X-A***	$u = \frac{X - 344.5}{30}$	***fu***
240-269	254.5	7	-90	-3	-21

270-299	284.5	19	-60	-2	-38
300-329	314.5	27	-30	-1	-27
330-359	344.5	15	0	0	0
360-389	374.5	12	30	1	12
390-419	404.5	12	60	2	24
420-449	434.5	8	90	3	24
	Total	100			-26

$$\bar{X} = A + h\frac{\sum_{i=1}^{n} f_i u_i}{N} = 344.5 - \frac{30*26}{100} = 336.7$$

Example: Following table gives the distribution of companies according to size of capital. Find the mean size of the capital of a company.

Capital(Lacs Rs.)	<5	<10	<15	<20	<25	<30
No. Of Companies	20	27	29	38	48	53

Solution: This is "less than"

class intervals	**Mid values (*X*)**	**Frequency (*f*)**	***u*=(X-12.5)/5**	***fu***
0-5	2.5	20	-2	-40
5-10	7.5	7	-1	-7
10-15	12.5	2	0	0
15-20	17.5	9	1	9
20-25	22.5	10	2	20
25-30	27.5	5	3	15
	Total	53		-3

$$\bar{X} = A + h\frac{\sum_{i=1}^{n} f_i u_i}{N} = 12.5 - \frac{5*3}{53} = 12.22 Lacs$$

1.7.2 WEIGHTED ARITHMETIC MEAN:

In the computation of simple arithmetic mean, equal importance is given to all the items. But this may not be so in all situations. If all the items are not of equal importance, then simple arithmetic mean will not be a good representative of the given data. Hence, weighing of different items becomes necessary. The weights are assigned to different items depending upon their importance, i.e., more important items are assigned more weight.

For example, to calculate mean wage of the workers of a factory, it would be wrong to compute simple arithmetic mean if there are a few workers (say managers) with very high wages while majority of the workers are at low level of wages. The simple arithmetic mean, in such a situation, will give a higher value that cannot be regarded as representative wage for the group. In order that the mean wage gives a realistic picture of the distribution, the wages of managers should be given less importance in its computation.

The mean calculated in this manner is called weighted arithmetic mean. The computation of weighted arithmetic is useful in many situations where different items are of unequal importance, e.g., the construction index numbers, computation of standardized death and birth rates, etc.

Hence the weighted arithmetic mean, μ, is given as:

$$\mu = \frac{\sum f_i \cdot w_i \cdot x_i}{\sum f_i \cdot w_i}$$

Where w_i is the weight for the i^{th} datum. f_i and x_i are defined same as those in the arithmetic mean for ungrouped and grouped data.

Example: From the Following result of two colleges *A* and *B*, Find out which of the two is better:

Examination	College A		College B	
	Appeared	Passed	Appeared	Passed
M.Sc.	60	40	200	160
M.A.	100	60	240	200
B.Sc.	200	150	200	140
B.A.	120	75	160	100

Solution : Performance of two colleges can be compared by taking weighted mean the two colleges.

Examination	College A			$w_A X_A$	College B			$w_B X_B$
	Appeared w_A	Passed	Pass Percentage X_A		Appeared w_B	Passed	Pass Percentage X_B	
M.Sc.	60	40	66.67	4000.2	200	160	80	16000
M.A.	100	60	60.00	6000.0	240	200	83.33	19999.2
B.Sc.	200	150	75.00	15000.0	200	140	70	14000
B.A.	120	75	62.50	7500.0	160	100	62.5	10000
Total	480	325		32500.2	800	600		59999.2

$$\overline{X} = \frac{\sum w_A.X_A}{\sum w_A} = \frac{32500.2}{480} = 67.71\%$$

$$\overline{X_w} = \frac{\sum w_B.X_B}{\sum w_B} = \frac{59999.2}{800} = 75\%$$

Since the weighted mean of Collage B is Higher , hence College B is Better.

1.8 Properties of Arithmetic Mean :

1. If all the observations assumed by a variable are constants, say "k", then arithmetic mean is also "k".
2. The algebraic sum of deviations of a set of observations from their arithmetic mean is zero.

That is, for unclassified data, $\sum(x - \bar{x}) = 0$. And for a grouped frequency distribution, $\sum f(x - \bar{x}) = 0$.

3. If there are two groups containing n_1 and n_2 observations $\bar{x}_1$ and $\bar{x}_2$ are the respective arithmetic means, then the combined arithmetic mean is given by $\overline{X} = \dfrac{n_1\overline{X}_1 + n_2\overline{X}_2}{n_1 + n_2}$

Example : If there are 130 teachers and 100 non-teaching employees in a college. The respective distributions of their monthly salaries are given in the following table:

Teachers		**Non- Teaching employee**	
Monthly Salary (Rs.)	**Frequency**	**Monthly Salary (Rs.)**	**Frequency**
4000-5000	10	1000-2000	21
5000-6000	16	2000-3000	45
6000-7000	22	3000-4000	28
7000-8000	67	4000-5000	06
8000-9000	15		
Total	**130**	**Total**	**100**

From the above data find:

(i) Average monthly salary of a teacher

(ii) Average monthly salary of non-teaching employee

(iii) Average monthly salary of employee(Teaching and Non Teaching)

Solution:

Teachers					**Non- Teaching employee**				
Monthly Salary (Rs.)	**Frequency**	**Mid values X**	$u = \frac{X-6500}{1000}$	***fu***	**Monthly Salary (Rs.)**	**Frequency**	**Mid values X**	$v = \frac{X-2500}{1000}$	***fv***
4000-5000	10	4500	-2	-20	1000-2000	21	1500	-1	-21
5000-6000	16	5500	-1	-16	2000-3000	45	2500	0	0
6000-7000	22	6500	0	0	3000-4000	28	3500	1	28
7000-8000	67	7500	1	67	4000-5000	06	4500	2	12
8000-9000	15	8500	2	30					
Total	**130**			**61**	**Total**	**100**			**19**

(i) Average monthly salary of a teacher $\overline{X}_1 = 6500 + 469.23 = \text{Rs.}6969.23$

(ii) Average monthly salary of nonteaching $\overline{X}_2 = 2500 + 190 = \text{Rs. } 2690$

(iii) Combined average salary $\overline{X} = \frac{n_1\overline{X}_1 + n_2\overline{X}_2}{n_1+n_2} = \frac{130*6969.23+100*2690}{230} = Rs.5108.70$

Example : The average rainfall for a week , excluding Sunday, was 10 cms. Deue to heavy rainfall on Sunday, the average for the week rose to 15 cms. How much rainfall was on Sunday?

Solution: A week can be trated as composed of two groups :

First consisting of 6 days excluding Sunday for which $N_1=6$ and $\overline{X}_1 = 10$.

The second group consisting of only Sunday for which $N_2=1$ and mean is $\overline{X}_2$ (Let).Now we have to determine the value of $\overline{X}_2$. Also given N=7 and $\overline{X} = 15$

Then $$\overline{X} = \frac{n_1\overline{X}_1 + n_2\overline{X}_2}{n_1+n_2} \Rightarrow 15 = \frac{6*10+1*\overline{X}_2}{7} \Rightarrow \overline{X}_2 = 45\,cms$$

Example : The mean age of the combined group of men and women is 30.5 years. If the mean age of the sub-group of men is 35 years and that of the sub-group of women is 25 years, find out percentage of men and women in the group.

Solution: Let x be the percentage of men in the combined group. Therefore, percentage of women = 100 - x.

We are given that $\overline{X}_1$(men) = 35 years and $\overline{X}_2$(women) = 25 years

Also $\overline{X}$(combined) = 30.5

$$30.5 = \frac{35x + 25\,(100-x)}{x+100-x} = \frac{35x+2500-25x}{100} \text{ or } 3050 = 10x + 2500$$

$$\Rightarrow x = \frac{550}{10} = 55\%.$$ Thus, there are 55% men and 45% women in the group.

To find missing frequency or a missing value

Example : The following is the distribution of weights (in lbs.) of 60 students of a class:

Weights	:	93 - 97	98 - 102	103 - 107	108 - 112	113 - 117
No. of Students	:	2	5	12	?	14
Weights	:	118 - 122	123 - 127	128 - 132	*Total*	
No. of Students	:	?	3	1	60	

If the mean weight of the students is 110.917, find the missing frequencies.

Solution: Let f_1 be the frequency of the class 108-112. Then, the frequency of the class 118-122 is given by 60 - (2 + 5 + 12 + 14 + 3 + 1 + f_1) = 23 - f_1

Writing this information in tabular form we have :

Weights (in lbs.)	No. of Students (f)	Mid-Values (X)	$u=\frac{X-110}{5}$	fu
93-97	2	95	-3	-6
98-102	5	100	-2	-10
103-107	12	105	-1	-12
108-112	f_1	110	0	0
113-117	14	115	1	14
118-122	$23-f_1$	120	2	$46-2f_1$
123-127	3	125	3	9
128-132	1	130	4	4
Total	60			$45-2f_1$

Using the formula for A.M., we can write $110.917 = 110 + \frac{(45-2f_1)5}{60}$

or $11.004 = 45 - 2f_1$ or $2f_1 = 33.996 = 34$ (approximately)

Thus, $f_1 = 17$ is the frequency of the class 108 - 112 and 23 - 17= 6 is the frequency of the class 118 - 122.

Example : Find out the missing item (x) of the following frequency distribution whose arithmetic mean is 11.37.

X :	5	7	(x)	11	13	16	20
f :	2	4	29	54	11	8	4

$$\overline{X} = \frac{\sum fX}{\sum f} = \frac{(5\times 2)+(7\times 4)+29x+(11\times 54)+(13\times 11)+(16\times 8)+(20\times 4)}{112}$$

$$11.37 = \frac{10+28+29x+594+143+128+80}{112} \text{ or } 11.37\times 112 = 983 + 29x$$

$$\therefore\ x = \frac{290.44}{29} = 10.015 = 10 \text{ (approximately)}$$

Example The arithmetic mean of 50 items of a series was calculated by a student as 20. However, it was later discovered that an item 25 was misread as 35. Find the correct value of mean.

Solution: $N = 50$ and $\overline{X} = 20$ $\therefore$ $\Sigma X_i = 50 \times 20 = 1000$

Thus $\Sigma X_{i\ (corrected)} = 1000 + 25 - 35 = 990$ and $\overline{X}_{(corrected)} = \frac{990}{50} = 19.8$

Alternatively, using property 7 :

$$\overline{X}_{new} = \overline{X}_{old} + \text{ average change in magnitude } = 20 - \frac{10}{50} = 20 - 0.2 = 19.8$$

Example The arithmetic mean of 50 items of a series was calculated by a student as 20. However, it was later discovered that an item 25 was misread as 35. Find the correct value of mean.

Solution: $N = 50$ and $\overline{X} = 20 \quad \therefore \quad \Sigma X_i = 50 \times 20 = 1000$

$$\text{Thus } \Sigma X_{i\,(\text{corrected})} = 1000 + 25 - 35 = 990 \text{ and } \overline{X}_{(\text{corrected})} = \frac{990}{50} = 19.8$$

Alternatively, using property 7 :

$$\overline{X}_{new} = \overline{X}_{old} + \text{ average change in magnitude} = 20 - \frac{10}{50} = 20 - 0.2 = 19.8$$

Example The sales of a balloon seller on seven days of a week are as given below:

Days	*Mon*	*Tue*	*Wed*	*Thu*	*Fri*	*Sat*	*Sun*
Sales (in Rs)	100	150	125	140	160	200	250

If the profit is 20% of sales, find his average profit per day.

Solution: Let P denote profit and S denote sales, $\therefore P = \frac{20}{100} \times S$

Using property 6, we can write $\overline{P} = \frac{20}{100} \times \overline{S}$ or $\overline{P} = \frac{1}{5} \times \overline{S}$

$$\text{Now } \overline{S} = \frac{100 + 150 + 125 + 140 + 160 + 200 + 250}{7} = 160.71$$

$$\therefore \quad \overline{P} = \frac{160.71}{5} = \text{Rs } 32.14$$

Hence, the average profit of the balloon seller is Rs 32.14 per day.

Alternatively, we can find profit of each day and take mean of these values.

Days	*Mon*	*Tue*	*Wed*	*Thu*	*Fri*	*Sat*	*Sun*
Profit (in Rs)	20	30	25	28	32	40	50

$$\overline{P} = \frac{20 + 30 + 25 + 28 + 32 + 40 + 50}{7} = \text{Rs } 32.14$$

1.9 Merits and Demerits of Arithmetic Mean

➢ **Merits:**

- It is rigidly defined.
- It is easy to calculate and simple to follow.
- It is based on all the observations.
- It is determined for almost every kind of data.
- It is finite and not indefinite.
- It is readily put to algebraic treatment.
- It is least affected by fluctuations of sampling.

- **Demerits:**
  - The arithmetic mean is highly affected by extreme values.
  - It cannot average the ratios and percentages properly.
  - It is not an appropriate average for highly skewed distributions.
  - It cannot be computed accurately if any item is missing.
  - The mean sometimes does not coincide with any of the observed value.
- **When not to use mean?**

  The mean is generally an inappropriate measure of average for data that are measured on ordinal scales. Ordinal data are rated according to a category where a higher score indicates a higher or better rank than a lower score.

1.9 MEDIAN:

The **median** of a set of data values is the middle value of the data set when it has been arranged in ascending order. That is, from the smallest value to the highest value.
Median of distribution is that value of the variate which divides it into two equal parts. In terms of frequency curve, the ordinate drawn at median divides the area under the curve into two equal parts. Median is a positional average because its value depends upon the position of an item and not on its magnitude.

1.10 Determination of Median

(a) When individual observations are given

The following steps are involved in the determination of median:

(i) The given observations are arranged in either ascending or descending order of magnitude.

(ii) Given that there are n observations, the median is given by:

1. The size of n+1/2 th observations, when n is odd.
2. The mean of the sizes of n/2th and n+1/2 of observations, when n is even.

Example : Find median of the following observations: 20, 15, 25, 28, 18, 16, 30.

Solution: Writing the observations in ascending order, we get 15, 16, 18, 20, 25, 28, 30.

Since n = 7, i.e., odd, the median is the size of 7+1/2 th, i.e., 4th observation.

Hence, median, denoted by M_d = 20.

Note: The same value of M_d will be obtained by arranging the observations in descending order of magnitude.

Example : Find median of the data : 245, 230, 265, 236, 220, 250.

Solution: Arranging these observations in ascending order of magnitude, we get 220, 230, 236, 245, 250, 265. Here n = 6, i.e., even.

Median will be arithmetic mean of the size of 6/2th, i.e., 3rd and (6/2 +1)th, i.e., 4th observations.

Hence M_d =236+245/2=240.5

Remarks: Consider the observations: 13, 16, 16, 17, 17, 18, 19, 21, 23. On the basis of the method given above, their median is 17.

According to the above definition of median, "half (i.e., 50%) of the observations should be below 17 and half of the observations should be above 17". Here we may note that only 3 observations are below 17 and 4 observations are above it and hence, the definition of median given above is some what ambiguous. In order to avoid this ambiguity, the median of a distribution may also be defined in the following way:

Median of a distribution is that value of the variate such that at least half of the observations are less than or equal to it and at least half of the observations are greater than or equal to it.

Based on this definition, we find that there are 5 observations which are less than or equal to 17 and there are 6 observations which are greater than or equal to 17. Since n = 9, the numbers 5 and 6 are both more than half, i.e., 4.5. Thus, median of the distribution is 17.

Further, if the number of observations is even and the two middle most observations are not equal, e.g., if the observations are 2, 2, 5, 6, 7, 8, then there are 3 observations (n/2=3) which are less than or equal to 5 and there are 4 (i.e., more than half) observations which are greater than or equal to 5.

Further, there are 4 observations which are less than or equal to 6 and there are 3 observations which are greater than or equal to 6. Hence, both 5 and 6 satisfy the conditions of the new definition of median. In such a case, any value lying in the closed interval [5, 6] can be taken as median. By convention we take the middle value of the interval as median. Thus, median is 5+6/2= 5.5

(b) When ungrouped frequency distribution is given:

In this case, the data are already arranged in the order of magnitude. Here, cumulative frequency is computed and the median is determined in a manner similar to that of individual observations.

Example : Locate median of the following frequency distribution:

x	10	11	12	13	14	15	16
f	8	15	25	20	12	10	5

Solution :

x	10	11	12	13	14	15	16
f	8	15	25	20	12	10	5
c.f.	8	23	48	68	80	90	95

Here N = 95, which is odd. Thus, median is size of (95+1/2)th i.e.,48th observation. From the table 48th observation is 12, Therefore Md = 12.

Alternative Method: N/2 = 95/2 = 47.5 Looking at the frequency distribution we note that there are 48 observations which are less than or equal to 12 and there are 72 (i.e., 95 - 23) observations which are greater than or equal to 12. Hence, median is 12.

Example *:* Locate median of the following frequency distribution :

x	0	1	2	3	4	5	6	7
f	7	14	18	36	51	54	52	20

Solution*:*

x	0	1	2	3	4	5	6	7
f	7	14	18	36	51	54	52	20
c.f.	7	21	39	75	126	180	232	252

Here N = 252, i.e., even.

Now N/2= 252/2 = 126 and N/2+1 = 127

Median is the mean of the size of 126th and 127th observation. From the table we note that 126th observation is 4 and 127th observation is 5.

M_d = 4+5/2= 4.5

Alternative Method: Looking at the frequency distribution we note that there are 126 observations which are less than or equal to 4 and there are 252 - 75 = 177 observations which are greater than or equal to 4. Similarly, observation 5 also satisfies this criterion. Therefore, median = 4+5/2 = 4.5.

(c) When grouped frequency distribution is given (Interpolation formula):

The determination of median, in this case, will be explained with the help of the following example.

Example : Suppose we wish to find the median of the following frequency distribution.

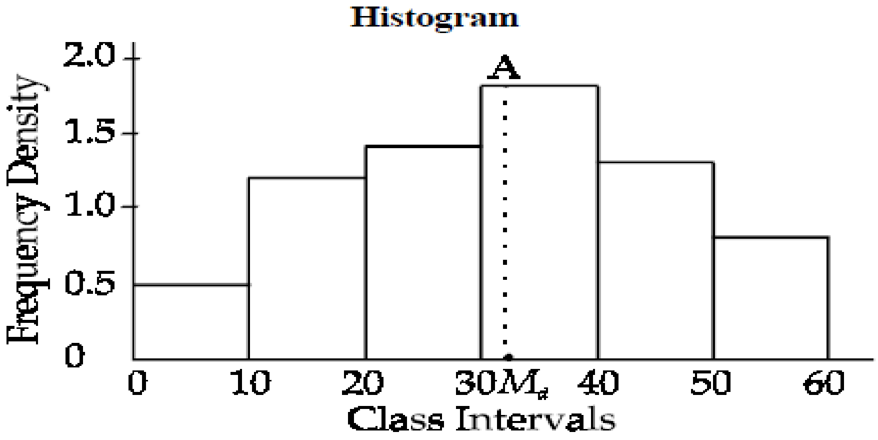
Classes	0-10	10-20	20-30	30-40	40-50	50-60
Freq.	5	12	14	18	13	8

Solution: The median of a distribution is that value of the variate which divides the distribution into two equal parts. In case of a grouped frequency distribution, this implies that the ordinate drawn at the median divides the area under the histogram into two equal parts. Writing the given data in a tabular form, we have:

Class Interval	**frequency**	**"less than" freq.**	**frequency density**
0-10	5	5	0.5
10-20	12	17	1.2
20-30	14	31	1.4
30-40	18	49	1.8
40-50	13	62	1.3
50-60	8	70	0.8

Frequency density in a class= frequency of the class/ width of the class = *f/h*

For the location of median, we make a histogram with heights of different rectangles equal to frequency density of the corresponding class. Such a histogram is shown below:



Since the ordinate at median divides the total area under the histogram into two equal parts, therefore we have to find a point (like Md as shown in the figure) on X - axis such that an ordinate (AMd) drawn at it divides the total area under the histogram into two equal parts.

We may note here that area under each rectangle is equal to the frequency of the corresponding class. Since area = length ´ breadth = frequency density× width of class = *f*/*h*× h = f.
Thus, the total area under the histogram is equal to total frequency N. In the given example N = 70, therefore *N*/2= 35. We note that area of first three rectangles is 5 + 12 + 14 = 31 and the area of first four rectangles is 5 + 12 + 14 + 18 = 49. Thus, median lies in the fourth class interval which is also termed as median class. Let the point, in median class, at which median lies be denoted by Md.
The position of this point should be such that the ordinate AMd (in the above histogram) divides the area of median rectangle so that there are only 35 - 31 = 4 observations to its left. From the histogram, we can also say that the position of Md should be such that

M_d-30/40-30= 4/18 Thus, M_d=40/18 + 30 =32.2

Writing the above equation in general notations, we have

$$median = l_1 + \frac{\frac{n}{2} - C}{f_m}(l_2 - l_1) = l_m + \frac{\frac{n}{2} - C}{f_m}(l_2 - l_1)$$

where: l_1 = lower class boundary of the median class;
n = total frequency;
C = cumulative frequency just before the median class;
f_m = frequency of the median;
l_2 = upper class boundary containing the median.

Remarks:1. Since the variable, in a grouped frequency distribution, is assumed to be continuous we always take exact value of *N*/2, including figures after decimals, when N is odd. 2. The above formula is also applicable when classes are of unequal width.

3. Median can be computed even if there are open end classes because here we need to know only the frequencies of classes preceding or following the median class.

(d) Determination of Median When 'greater than' type cumulative frequencies are given:

By looking at the histogram, we note that one has to find a point denoted by Md such that area to the right of the ordinate at Md is 35. The area of the last two rectangles is 13 + 8 = 21. Therefore, we have to get 35 - 21 = 14 units of area from the median rectangle towards right of the ordinate. Let Um be the upper limit of the median class. Then the formula for median in this case can be written as

$$median = l_2 - \frac{\frac{n}{2} - C}{f_m}(l_2 - l_1) = u_m - \frac{\frac{n}{2} - C}{f_m}(l_2 - l_1)$$

Note that C denotes the 'greater than type' cumulative frequency of classes following the median class. Applying this formula to the above example, we get M_d=40-(35-21)/18 x 10 = 32.2

Example: Calculate median of the following data :

Height (Inch)	3-4	4-5	5-6	6-7	7-8	8-9	9-10	10-11
No. of Saplings	3	7	12	16	22	20	13	7

Solution :

Height (Inch)	No. of Saplings	"less than" type c.f.
3-4	3	3
4-5	7	10
5-6	12	22
6-7	16	38
7-8	22	60
8-9	20	80
9-10	13	93
10-11	7	100

Since $N/2$=100/2= 50, the median class is 7- 8. Further, Lm = 7, h = 1, fm = 22 and C = 38.

Thus, M_d = 7 +50-38/22x1 = 7.55 inches

Example*:* The following table gives the distribution of marks by 500 students in an examination. Obtain median of the given data.

Marks	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79
No. Of students	30	40	50	48	24	162	132	14

Solution*:* Since the class intervals are inclusive, therefore, it is necessary to convert them into class boundaries.

Class Intervals	Class Boundaries	Frequency	'Less than' type c.f.
0-9	−0.5-9.5	30	30
10-19	9.5-19.5	40	70
20-29	19.5-29.5	50	120
30-39	29.5-39.5	48	168
40-49	39.5-49.5	24	192
50-59	49.5-59.5	162	354
60-69	59.5-69.5	132	486
70-79	69.5-79.5	14	500

Since N/2 = 250, the median class is 49.5 - 59.5 and, therefore, Lm = 49.5, h = 10, fm= 162, C=192 .

Thus M_d= 49.5 + 250-192/162 x10 = 53.08 marks

Example: The weekly wages of 1,000 workers of a factory are shown in the following table. Calculate median.

Weekly Wages (less than)	:	425	475	525	575	625	675	725	775	825	875
No. of Workers	:	2	10	43	123	293	506	719	864	955	1000

Solution: The above is a 'less than' type frequency distribution. This will first be converted into class intervals

Class Intervals	Frequency	Less than c.f.
less than 425	2	2
425 - 475	8	10
475 - 525	33	43
525 - 575	80	123
575 - 625	170	293
625 - 675	213	506
675 - 725	213	719
725 - 775	145	864
775 - 825	91	955
825 - 875	45	1000

Since *N/2*= 500, the median class is 625 - 675. On substituting various values in the formula for median, we get

Age greater than (in yrs)	:	0	10	20	30	40	50	60	70
No. of Persons	:	230	218	200	165	123	73	28	8

Example : Find the median of the following data:

Class Interval	Grater than c.f.	frequency
0-10	230	12
10-20	218	18
20-30	200	35
30-40	165	42
40-50	123	50
50-60	73	45
60-70	28	20

70 and above	8	8

Solution: Note that it is 'greater than' type frequency distribution

$$median = l_2 - \frac{\frac{n}{2} - C}{f_m}(l_2 - l_1) = u_m - -\frac{\frac{n}{2} - C}{f_m}(l_2 - l_1) = 50 - \frac{115 - 73}{50}(10) = 41.6\, years$$

Since $N/2$=230/2= 115, the median class is 40 - 50.

Example : The following table gives the daily profits (in Rs) of 195 shops of a town. Calculate mean and median.

Profit	50-60	60-70	70-80	80-90	90-100	100-110	110-120	120-130	130-140
No. Of Shops	15	20	32	35	33	22	50	10	8

Solution:

Profit	No. Of Shops (f)	Mid Value(X)	u=X-95 /10	fu	less than c.f.
50-60	15	55	-4	-60	15
60-70	20	65	-3	-60	35
70-80	32	75	-2	-64	67
80-90	35	85	-1	-35	102
90-100	33	95	0	0	135
100-110	22	105	1	22	157
110-120	50	115	2	40	177
120-130	10	125	3	30	187
130-140	8	135	4	32	195
Total	195			-95	

$$\overline{X} = A + \frac{\sum fu}{N} \times h = 95 - \frac{95}{195} \times 10 = Rs.90.13$$

Since N/2= 195/2=97.5 hence median class is 80-90.

$$M_d = 80 + \frac{97.5 - 67}{35} \times 10 = Rs.88.71$$

Example : Find median of the following distribution:

Mid - Values	:	1500	2500	3500	4500	5500	6500	7500
Frequency	:	27	32	65	78	58	32	8

Solution: Since the mid-values are equally spaced, the difference between their two successive values will be the width of each class interval. This width is 1,000. On subtracting and adding half of this, i.e., 500 to each of the mid-values, we get the lower and the upper limits of the respective class intervals. After this, the calculation of median can be done in the usual way.

Mid - Values	Class Intervals	Frequency	c. f. (less than)
1500	1000 - 2000	27	27
2500	2000 - 3000	32	59
3500	3000 - 4000	65	124
4500	4000 - 5000	78	202
5500	5000 - 6000	58	260
6500	6000 - 7000	32	292
7500	7000 - 8000	8	300

1.11 Determination of Missing Frequencies :

If the frequencies of some classes are missing, however, the median of the distribution is known, and then these frequencies can be determined by the use of median formula.

Example: The following table gives the distribution of daily wages of 900 workers. However, the frequencies of the classes 40 - 50 and 60 - 70 are missing. If the median of the distribution is Rs 59.25, find the missing frequencies.

Wages (Rs)	:	30-40	40-50	50-60	60-70	70-80
No. of Workers	:	120	?	200	?	185

Solution: Let $f1$ and f2 be the frequencies of the classes 40 - 50 and 60 - 70 respectively.

Class Intervals	Frequency	c.f. (less than)
30-40	120	120
40-50	f_1	$120+f_1$
50-60	200	$320+f_1$
60-70	f_2	$320+f_1+f_2$
70-80	185	900

Since median is given as 59.25, the median class is 50 - 60.

Therefore, we can write $59.25 = 50 + \dfrac{450-(120+f_1)}{200} \times 10 \Rightarrow f_1 = 145$

Further f_2=900-(120+145+200+185)=250

Properties of Median

1. It is a positional average.
2. It can be shown that the sum of absolute deviations is minimum when taken from median. This property implies that median is centrally located.

1.12 Merits and Demerits of Median

(a) Merits *of Median*

- It is very easy to calculate and is readily understood.
- Median is not affected by the extreme values. It is independent of the range of series.
- Median can be measured graphically.
- Median serves as the most appropriate average to deal with qualitative data.
- Median value is always certain and specific value in the series.
- Median is often used to convey the typical observation. It is primarily affected by the number of observations rather than their size.

(b) Demerits *of* Median

- Median does not represent the measure of such series of which different values are wide apart from each other.
- Median is erratic if the number of items is small.
- Median is incapable of further algebraic treatment.
- Median is very much affected by the sampling fluctuations.
- It is affected much more by fluctuations of sampling than A.M.
- Median cannot be used for further algebraic treatment. Unlike mean we can neither find total of terms as in case of A.M. nor median of some groups when combined.
- In a continuous series it has to be interpolated. We can find its true-value only if the frequencies are uniformly spread over the whole class interval in which median lies.
- If the number of series is even, we can only make its estimate; as the A.M. of two middle terms is taken as Median.

1.13 Uses of Median

1. It is an appropriate measure of central tendency when the characteristics are not measurable but different items are capable of being ranked.
2. Median is used to convey the idea of a typical observation of the given data.
3. Median is the most suitable measure of central tendency when the frequency distribution is skewed. For example, income distribution of the people is generally positively skewed and median is the most suitable measure of average in this case.
4. Median is often computed when quick estimates of average are desired.
5. When the given data has class intervals with open ends, median is preferred as a measure of central tendency since it is not possible to calculate mean in this case.

1.14 When to use median?

The median is a good measure of the average value when the data include exceptionally high or low values because these have little influence on the outcome. The median is the most suitable measure of average for data classified on an ordinal scale.

1.15 MODE:

The concept of mode, as a measure of central tendency, is preferable to mean and median when it is desired to know the most typical value, e.g., the most common size of shoes, the most common size of a ready-made garment, the most common size of income, the most common size of pocket expenditure of a college student, the most common size of a family in a locality, the most common duration of cure of viral-fever, the most popular candidate in an election, etc.

i.e. Mode is that value of the variate which occurs maximum number of times in a distribution and around which other items are densely distributed.

In the words of ***Croxton and Cowden***, "The mode of a distribution is the value at the point around which the items tend to be most heavily concentrated. It may be regarded the most typical of a series of values." Further, according to A.M. Tuttle, "Mode is the value which has the greatest frequency density in its immediate neighbourhood."

If the frequency distribution is regular, then mode is determined by the value corresponding to maximum frequency. There may be a situation where concentration of observations around a value having maximum frequency is less than the concentration of observations around some other value. In such a situation, mode cannot be determined by the use of maximum frequency criterion. Further, there may be concentration of observations around more than one value of the variable and, accordingly, the distribution is said to be bimodal or multi-modal depending upon whether it is around two or more than two values.

1.19 Determination of Mode

(a) When data are either in the form of individual observations or in the form of ungrouped frequency distribution :

Given individual observations, these are first transformed into an ungrouped frequency distribution. The mode of an ungrouped frequency distribution can be determined in two ways, as given below:

1. By inspection or
2. By method of Grouping

(i) ***By inspection:*** When a frequency distribution is fairly regular, then mode is often determined by inspection. It is that value of the variate for which frequency is maximum. By a fairly regular frequency distribution we mean that as the values of the variable increase the corresponding frequencies of these values first increase in a gradual manner and reach a peak at certain value and, finally, start declining gradually in, approximately, the same manner as in case of increase.

Example: Compute mode of the following data:

3,4,5,10,15,3,6,7,9,12,10,16,18,20,10,9,8,19,11,14,10,13,17,9,11

Solution: Writing this in the form of a frequency distribution, we get

Value	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Freq.	2	1	1	1	1	1	3	4	2	1	1	1	1	1	1	1	1	1

Since the frequency of 10 is maximum(=4) hence mode is 10.

Remarks:

1. If the frequency of each possible value of the variable is same, there is no mode.
2. If there are two values having maximum frequency, the distribution is said to be bimodal.

Example *:* Compute mode of the following distribution

X	5	10	15	20	25	30	35	40
f	2	4	6	10	15	9	5	4

Solution*:* The given distribution is fairly regular. Therefore, the mode can be determined just by inspection. Since for X = 25 the frequency is maximum, mode = 25.

(ii) ***By method of Grouping:*** This method is used when the frequency distribution is not regular. Let us consider the following example to illustrate this method.

Example : Determine the mode of the following distribution.

x	10	11	12	13	14	15	16	17	18	19
f	8	15	20	100	98	95	90	75	50	30

Solution: This distribution is not regular because there is sudden increase in frequency from 20 to 100.

Therefore, mode cannot be located by inspection and hence the method of grouping is used. Various steps involved in this method are as follows:

1. Prepare a table consisting of 6 columns in addition to a column for various values of X.
2. In the first column, write the frequencies against various values of X as given in the question.
3. In second column, the sum of frequencies, starting from the top and grouped in twos, are written.
4. In third column, the sum of frequencies, starting from the second and grouped in twos, is written.
5. In fourth column, the sum of frequencies, starting from the top and grouped in threes is written.
6. In fifth column, the sum of frequencies, starting from the second and grouped in threes is written.

7. In the sixth column, the sum of frequencies, starting from the third and grouped in threes is written.

The highest frequency total in each of the six columns is identified and analyzed to determine mode. We apply this method for determining mode of the above example.

X	f (1)	(2)	(3)	(4)	(5)	(6)
10	8	23		43		
11	15		35		135	
12	20	120				218
13	(100)		(198)	(293)		
14	98	(193)			(283)	
15	95		185			(260)
16	90	165		215		
17	75		125		155	
18	50	80				
19	30					

Analysis Table

Columns	10	*V* 11	*A* 12	*R* 13	*I* 14	*A* 15	*B* 16	*L* 17	*E* 18	19
1				1						
2					1	1				
3				1	1					
4				1	1	1				
5					1	1	1			
6						1	1	1		
Total	0	0	0	3	4	4	2	1	0	0

Since the value 14 and 15 are both repeated maximum number of times in the analysis table, therefore, mode is ill defined. Mode in this case can be approximately located by the use of the following formula, which will be discussed later, in this chapter.

Mode = 3 Median - 2 mean

Calculation of Median and Mean

X	10	11	12	13	14	15	16	17	18	19	*Total*
f	8	15	20	100	98	95	90	75	50	30	581
$c.f.$	8	23	43	143	241	336	426	501	551	581	
fX	80	165	240	1300	1372	1425	1440	1275	900	570	8767

$$\text{Median} = \text{Size of } \left(\frac{581+1}{2}\right)\text{th, i.e., 291st observation} = 15.\ \text{Mean} = \frac{8767}{581} = 15.09$$

$$\therefore \quad \text{Mode} = 3 \times 15 - 2 \times 15.09 = 45 - 30.18 = 14.82$$

Remarks: If the most repeated values, in the above analysis table, were not adjacent, the distribution would have been bi-modal, i.e., having two modes

Example : From the following data regarding weights of 60 students of a class, find modal weight:

Weight	50	51	52	53	54	55	56	57	58	59	60
No. Of Students	2	4	5	6	8	5	4	7	11	5	3

Solution*:* Since the distribution is not regular, method of grouping will be used for determination of mode.

Grouping Table

Weight (in Kgs.)	*Frequency* (1)	(2)	(3)	(4)	(5)	(6)
50	2	6		11		
51	4		9		15	
52	5	11				**19**
53	6		14	19		
54	8	13			17	
55	5		9			16
56	4	11		**22**		
57	7		**18**		**23**	
58	**11**	**16**				**19**
59	5		8			
60	3					

Analysis Table

Columns	50	51	W 52	E 53	I 54	G 55	H 56	T 57	S 58	59	60
1									1		
2									1	1	
3								1	1		
4							1	1	1		
5								1	1	1	
6			1	1	1				1	1	1
Total	0	0	1	1	1	0	1	3	6	3	1

Since the value 58 has occurred maximum number of times, therefore, mode of the distribution is 58 kgs.

(b) When data are in the form of a grouped frequency distribution

The following steps are involved in the computation of mode from a grouped frequency distribution.

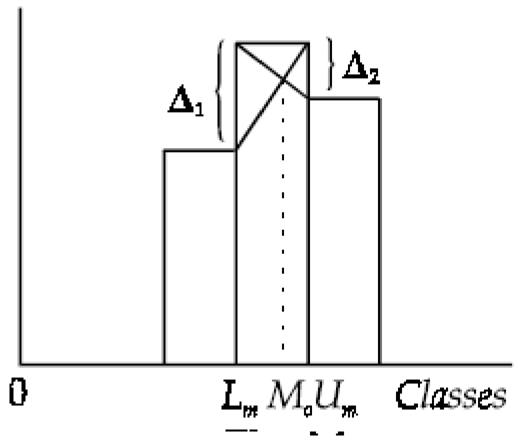
(i) ***Determination of modal class:*** It is the class in which mode of the distribution lies. If the distribution is regular, the modal class can be determined by inspection, otherwise, by method of grouping.

Exact location of mode in a modal class (interpolation formula): The exact location of mode, in a modal class, will depend upon the frequencies of the classes immediately preceding and following it. If these frequencies are equal, the mode would lie at the middle of the modal class interval. However, the position of mode would be to the left or to the right of the middle point depending upon whether the frequency of preceding class is greater or less than the frequency of the class following it. The exact location of mode can be done by the use of interpolation formula, developed below:

Let the modal class be denoted by Lm - Um, where Lm and Um denote its lower and the upper limits respectively. Further, let fm be its frequency and h its width. Also let f1 and f2 be the respective frequencies of the immediately preceding and following classes.

We assume that the width of all the class intervals of the distribution is equal. If these are not equal, make them so by regrouping under the assumption that frequencies in a class are uniformly distributed.

Make a histogram of the frequency distribution with height of each rectangle equal to the frequency of the corresponding class. Only three rectangles, out of the complete histogram, that are necessary for the purpose are shown in the above figure.



Let $\Delta_1 = f_m - f_1$ and $\Delta_2 = f_m - f_2$. Then the mode, denoted by M_o, will divide the modal class interval in the ratio $\frac{\Delta_1}{\Delta_2}$. The graphical location of mode is shown in Fig.

To derive a formula for mode, the point M_o in the figure, should be such that

$$\frac{M_o - L_m}{U_m - M_o} = \frac{\Delta_1}{\Delta_2} \text{ or } M_o\Delta_2 - L_m\Delta_2 = U_m\Delta_1 - M_o\Delta_1$$

$$\Rightarrow (\Delta_1 + \Delta_2)M_o = L_m\Delta_2 + U_m\Delta_1 = L_m\Delta_2 + (L_m + h)\Delta_1 \quad (\text{where } U_m = L_m + h)$$

$$= (\Delta_1 + \Delta_2) L_m + \Delta_1 h$$

Dividing both sides by $\Delta_1 + \Delta_2$, we have

$$M_o = L_m + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times h \qquad \text{.... (1)}$$

By slight adjustment, the above formula can also be written in terms of the upper limit (U_m) of the modal class.

$$M_o = U_m - h + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times h = U_m - \left[1 - \frac{\Delta_1}{\Delta_1 + \Delta_2}\right] \times h$$

$$= U_m - \left[\frac{\Delta_2}{\Delta_1 + \Delta_2} \times h\right] \qquad \text{.... (2)}$$

Replacing Δ_1 by $f_m - f_1$ and Δ_2 by $f_m - f_2$, the above equations can be written as

$$M_o = L_m + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times h \text{ and } M_o = U_m - \frac{f_m - f_2}{2f_m - f_1 - f_2} \times h$$

Note: The above formulae are applicable only to a unimodal frequency distribution.

Example : The monthly profits (in Rs) of 100 shops are distributed as follows:

Profit per shop	0-100	100-200	200-300	300-400	400-500	500-600
No. Of Shopes	12	18	27	20	17	6

Determine the 'modal value' of the distribution graphically and verify the result by calculation.

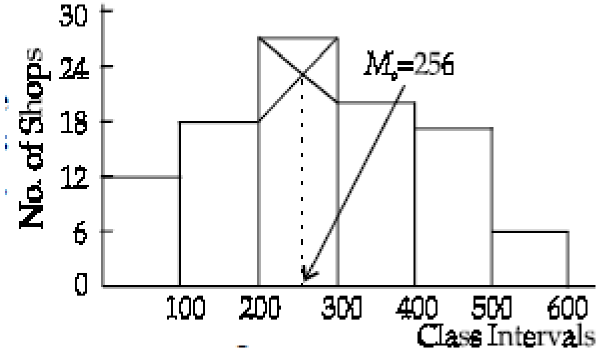
Solution: Since the distribution is regular, the modal class would be a class having the highest frequency. The modal class, of the given distribution, is 200 - 300.

Graphical Location of Mode

To locate mode we draw a histogram of the given frequency distribution. The mode is located as shown in figure. From the figure, mode = Rs 256.

Determination of Mode by interpolation formula

Since the modal class is 200 - 300, Lm = 200, D1 = 27 - 18 = 9, D2 = 27 - 20 = 7 and h = 100.



$$\therefore \quad M_o = 200 + \frac{9}{9+7} \times 100 = \text{Rs } 256.25$$

Example : The frequency distribution of marks obtained by 60 students of a class in a college is given below :

Marks	:	30 - 34	35 - 39	40 - 44	45 - 49	50 - 54	55 - 59	60 - 64
Frequency	:	3	5	12	18	14	6	2

Find mode of the distribution.

Solution: The given class intervals are first converted into class boundaries, as given in the following table :

Marks	:	29.5 - 34.5	34.5 - 39.5	39.5 - 44.5	44.5 - 49.5
Frequency	:	3	5	12	18
Marks	:	49.5 - 54.5	54.5 - 59.5	59.5 - 64.5	
Frequency	:	14	6	2	

We note that the distribution is regular. Thus, the modal class, by inspection, is 44.5 - 49.5.

Further, $L_m = 44.5$, $\Delta_1 = 18 - 12 = 6$, $\Delta_2 = 18 - 14 = 4$ and $h = 5$

$$\therefore \quad \text{Mode} = 44.5 + \frac{6}{6+4} \times 5 = 47.5 \text{ marks}$$

Example : Calculate mode of the following data :

Weekly Wages (Rs)	:	200-250	250-300	300-350	350-400
No. of Workers	:	4	6	20	12
Weekly Wages (Rs)	:	400-450	450-500	500-550	550-600
No. of Workers	:	33	17	8	2

Solution: Since the frequency distribution is not regular, the modal class will be determined by the method of grouping.

Grouping Table

Weekly wages (in Rs)	*f* (1)	(2)	(3)	(4)	(5)	(6)
200 - 250	4	10		30		
250 - 300	6		26		38	
300 - 350	20	32				**65**
350 - 400	12		**45**	**62**		
400 - 450	**33**	**50**			**58**	
450 - 500	17		25			27
500 - 550	8	10				
550 - 600	2					

Analysis Table

Columns	300 - 350	350 - 400	400 - 450	450 - 500	500 - 550
1			1		
2			1	1	
3		1	1		
4		1	1	1	
5			1	1	1
6	1	1	1		
Total	1	3	6	3	1

The modal class, from analysis table, is 400 - 500.

Thus, $L_m = 400$, $\Delta_1 = 33 - 12 = 21$, $\Delta_2 = 33 - 17 = 16$ and $h = 50$

$$\text{Hence, mode} = 400 + \frac{21}{37} \times 50 = \text{Rs } 428.38$$

Example Calculate mode of the following distribution :

Weights (lbs.)	:	*below* 100	*below* 110	*below* 120	*below* 130	*below* 140
No. of Students	:	4	6	24	46	67
Weights (lbs.)	:	*below* 150	*below* 160	*below* 170	*below* 180	
No. of Students	:	86	96	99	100	

Solution: Rewriting the above distribution in the form of a frequency distribution with class limits, we get

Weights (lbs.)	:	*Less than* 100	100 - 110	110 - 120	120 - 130	130 - 140
Frequency	:	4	2	18	22	21
Weights (lbs.)	:	140 - 150	150 - 160	160 - 170	170 - 180	
Frequency	:	19	10	3	1	

We note that there is a concentration of observations in classes 120 - 130 and 130 - 140, therefore, modal class can be determined by the method of grouping.

Grouping Table

Weights (in lbs.)	*f* (1)	(2)	(3)	(4)	(5)	(6)
less than 100	4	6		24		
100 - 110	2		20		42	
110 - 120	18	40				61
120 - 130	22		43	62		
130 - 140	21	40			50	
140 - 150	19		29			32
150 - 160	10	13		14		
160 - 170	3		4			
170 - 180	1					

Analysis Table

Columns	110 - 120	120 - 130	130 - 140	140 - 150	150 - 160
1		1			
2	1	1	1	1	
3		1	1		
4		1	1	1	
5			1	1	1
6	1	1	1		
Total	2	5	5	3	1

Since the two classes, 120 - 130 and 130 - 140, are repeated maximum number of times in the above table, it is not possible to locate modal class even by the method of grouping. However, an approximate value of mode is given by the empirical formula:

Mode = 3 Median - 2 Mean

Looking at the cumulative frequency column, given in the question, the median class is 130 - 140. Thus, Lm = 130, C = 46, fm = 21, h = 10.

∴ Md = 130 + 50-46/21 x 10 = 131.9 lbs.

Assuming that the width of the first class is equal to the width of second, we can write

Mid - Values (X)	95	105	115	125	135	145	155	165	175	*Total*
f	4	2	18	22	21	19	10	3	1	100
$u \quad \frac{X \quad 135}{10}$	4	3	2	1	0	1	2	3	4	
fu	16	6	36	22	0	19	20	9	4	28

Thus, $\overline{X} = 135 - \frac{28 \times 10}{100} = 135 - 2.8 = 132.2$ lbs.

Using the values of mean and median, we get

$$M_{\_} = 3 \times 131.9 - 2 \times 132.2 = 131.3 \text{ lbs.}$$

Remarks: Another situation, in which we can use the empirical formula, rather than the interpolation formula, is when there is maximum frequency either in the first or in the last class.

Calculation of Mode when either D1 or D2 is negative:

The interpolation formula, for the calculation of mode, is applicable only if both D1 and D2 are positive. If either D1 or D2 is negative, we use an alternative formula that gives only an approximate value of the mode.

We recall that the position of mode, in a modal class, depends upon the frequencies of its preceding and following classes, denoted by f1 and f2 respectively. If f1 = f2, the mode will be at the middle point which can be obtained by adding $f_2/(f_1+f_2)*h$ to the lower limit of the modal class or, equivalently, it can be obtained by subtracting $f_2/(f_1+f_2)*h$ from its upper limit. We may note that $f_1/(f_1+f_2)=f_2/(f_1+f_2)= 1/2$ when $f1 = f2$.

Further, if f2 > f1, the mode will lie to the right of the mid-value of modal class and, therefore, the ratio $f2/f1$ $f2$ will be greater than 1/2 . Similarly, if $f2 < f1$, the mode will lie to the left of the mid-value of modal class and, therefore, the ratio $f2/f1$ $f2$ will be less than ½ . Thus, we can write an alternative formula for mode as:

$$\text{Mode} = L_m + \frac{f_2}{f_1 + f_2} \times h \quad \text{or equivalently,} \qquad \text{Mode} = U_m - \frac{f_2}{f_1 + f_2} \times h$$

Remarks: The above formula gives only an approximate estimate of mode vis-a-vis the interpolation formula.

Example : Calculate mode of the following distribution.

Mid - Values	:	5	15	25	35	45	55	65	75
Frequency	:	7	15	18	30	31	4	3	1

Solution: The mid-values with equal gaps are given, therefore, the corresponding class intervals would be 0 - 10, 10 - 20, 20 - 30, etc.

Since the given frequency distribution is not regular, the modal class will be determined by the method of grouping.

Grouping Table

Class Intervals	f (1)	(2)	(3)	(4)	(5)	(6)
0 - 10	7	22		40		
10 - 20	15		33		63	
20 - 30	18	48				79
30 - 40	30		61	65		
40 - 50	31	35			38	
50 - 60	4		7			8
60 - 70	3	4				
70 - 80	1					

Analysis Table

Columns	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60
1				1	
2		1	1		
3			1	1	
4			1	1	1
5	1	1	1		
6		1	1	1	
Total	1	3	5	4	1

From the analysis table, the modal class is 30 - 40.

Therefore, $L_m = 30$, $\Delta_1 = 30 - 18 = 12$, $\Delta_2 = 30 - 31 = -1$ (negative) and $h = 10$.

We note that the interpolation formula is not applicable.

$$\text{Mode} = L_m + \frac{f_2}{f_1 + f_2} \times h = 30 + \frac{31}{18+31} \times 10 = 36.33$$

Example : The rate of sales tax as a percentage of sales, paid by 400 shopkeepers of a market during an assessment year ranged from 0 to 25%. The sales tax paid by 18% of them was not greater than 5%. The median rate of sales tax was 10% and 75th percentile rate of sales tax was 15%. If only 8% of the shopkeepers paid sales tax at a rate greater than 20% but not greater than 25%, summarize the information in the form of a frequency distribution taking intervals of 5%. Also find the modal rate of sales tax.

Solution: The above information can be written in the form of the following distribution:

Class Intervals (in percentage)	*No. of Shopkeepers*
0-5	$\frac{18}{100} \times 400 = 72$
5-10	$200 - 72 = 128$
10-15	$300 - 200 = 100$
15-20	$400 - 72 - 128 - 100 - 32 = 68$
20-25	$\frac{8}{100} \times 400 = 32$

By inspection, the modal class is 5 - 10.

$$\therefore \qquad M_o = 5 + \frac{128 - 72}{128 - 72 + 128 - 100} \times 5 = 8.33\%$$

Merits and Demerits of Mode

Merits

1. Mode is very simple measure of central tendency. Sometimes, just at the series is enough to locate the model value. Because of its simplicity, it s a very popular measure of the central tendency.
2. Compared top mean, mode is less affected by marginal values in the series. Mode is determined only by the value with highest frequencies.
3. Mode can be located graphically, with the help of histogram.
4. Mode is that value which occurs most frequently in the series. Accordingly, mode is the best representative value of the series.
5. The calculation of mode does not require knowledge of all the items and frequencies of a distribution. In simple series, it is enough if one knows the items with highest frequencies in the distribution.

Demerits

Mode is an uncertain and vague measure of the central tendency.

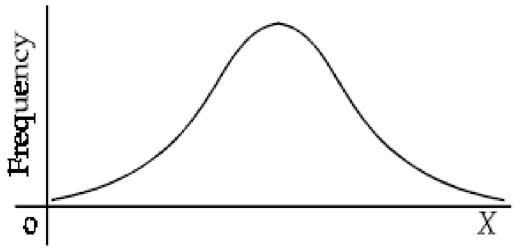
- Unlike mean, mode is not capable of further algebraic treatment.
- With frequencies of all items are identical, it is difficult to identify the modal value.
- Calculation of mode involves cumbersome procedure of grouping the data. If the extent of grouping changes there will be a change in the model value.
- It ignores extreme marginal frequencies. To that extent model value is not a representative value of all the items in a series.

When to use the mode ?

The mode is the measure of average that can be used with nominal data. For example, late-night users of the library were classified by faculty as: 14% science students, 32% social science students, and 54% biological sciences students. No median or mean can be calculated but the mode is biological science students as students from this faculty were the most common.

1.8 RELATION BETWEEN MEAN MEDIAN AND MODE :

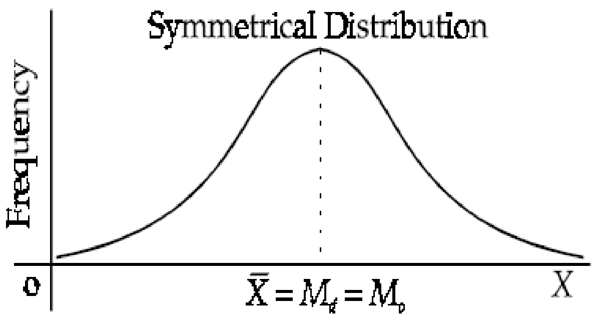
The relationship between the above measures of central tendency will be interpreted in terms of a continuous frequency curve. If the number of observations of a frequency distribution is increased gradually, then accordingly, we need to have more number of classes, for approximately the same range of values of the variable, and simultaneously, the width of the corresponding classes would decrease. Consequently, the histogram of the frequency distribution will get transformed into a smooth frequency curve, as shown in the following figure.



For a given distribution, the mean is the value of the variable which is the point of balance or centre of gravity of the distribution. The median is the value such that half of the observations are below it and remaining half are above it. In terms of the frequency curve, the total area under the curve is divided into two equal parts by the ordinate at median. Mode of a distribution is a value around which there is maximum concentration of observations and is given by the point at which peak of the curve occurs.

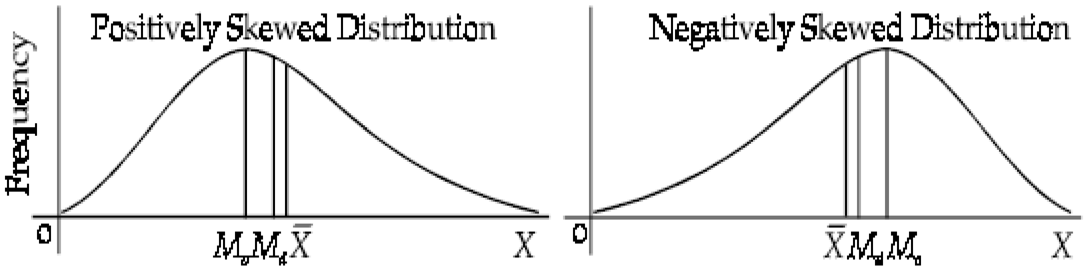
Mean Median Mode relationship in Symmetric distribution

For a symmetrical distribution, all the three measures of central tendency are equal i.e. $\boldsymbol{X = M_d = M_o}$, as shown in the following figure.



Imagine a situation in which the symmetrical distribution is made asymmetrical or positively (or negatively) skewed by adding some observations of very high (or very low) magnitudes, so that the right hand (or the left hand) tail of the frequency curve gets elongated.

Consequently, the three measures will depart from each other. Since mean takes into account the magnitudes of observations, it would be highly affected. Further, since the total number of observations will also increase, the median would also be affected but to a lesser extent than mean. Finally, there would be no change in the position of mode. More specifically, we shall have Mo < Md < $\bar{X}$, when skewness is positive and $\bar{X}$ < Md < Mo, when skewness is negative, as shown in the following figure.



1.9 Empirical Relation between Mean, Median and Mode

Empirically, it has been observed that for a moderately skewed distribution, the difference between mean and mode is approximately three times the difference between mean and median,

$$\text{i.e., } \bar{X} - M_o = 3\left(\bar{X} - M_d\right)$$

This relation can be used to estimate the value of one of the measures when the values of the other two are known.

Example : (a) The mean and median of a moderately skewed distribution are 42.2 and 41.9 respectively. Find mode of the distribution.

(b) For a moderately skewed distribution, the median price of men's shoes is Rs 380 and modal price is Rs 350. Calculate mean price of shoes.

Solution: (a) Here, mode will be determined by the use of empirical formula.

$$\bar{X} - M_o = 3(\bar{X} - M_d) \quad or \quad M_o = 3M_d - 2\bar{X}$$

It is given that $\bar{X} = 42.2$ and $M_d = 41.9$

$\therefore \qquad M_o = 3 \times 41.9 - 2 \times 42.2 = 125.7 - 84.4 = 41.3$

(b) Using the empirical relation, we can write $\bar{X} = \dfrac{3M_d - M_o}{2}$

It is given that M_d = Rs 380 and M_o = Rs. 350

$$\therefore \qquad \bar{X} = \frac{3 \times 380 - 350}{2} = \text{Rs } 395$$

(b) Using the empirical relation, we can write $\bar{X} = \dfrac{3M_d - M_o}{2}$

It is given that M_d = Rs 380 and M_o = Rs. 350

$$\therefore \qquad \bar{X} = \frac{3 \times 380 - 350}{2} = \text{Rs } 395$$

$$\therefore \qquad M_d = 0 + \frac{44.5 - 0}{45} \times 10 = 9.89$$

Also $\quad \bar{X} = 25 - \dfrac{97 \times 10}{89} = 14.10$

Thus, $M_o = 3M_d - 2\bar{X} = 3 \times 9.89 - 2 \times 14.10 = 1.47$

Example : Find mode of the following distribution :

Class Intervals	:	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50
Frequency	:	45	20	14	7	3

Solution: Since the highest frequency occurs in the first class interval, the interpolation formula is not applicable. Thus, mode will be calculated by the use of empirical formula.

Calculation of Mean and Median

Class Intervals	*Frequency*	*c. f.*	*Mid - Values*	*u* $\frac{X\ \ 25}{10}$	*fu*
0 - 10	45	45	5	2	90
10 - 20	20	65	15	1	20
20 - 30	14	79	25	0	0
30 - 40	7	86	35	1	7
40 - 50	3	89	45	2	6
Total	89				97

Since $\dfrac{N}{2} = \dfrac{89}{2} = 44.5$, the median class is 0 - 10.

$\therefore \qquad M_d = 0 + \dfrac{44.5 - 0}{45} \times 10 = 9.89$

Also $\quad \overline{X} = 25 - \dfrac{97 \times 10}{89} = 14.10$

Thus, $M_o = 3M_d - 2\overline{X} = 3 \times 9.89 - 2 \times 14.10 = 1.47$

Example : Estimate mode of the following distribution :

Weekly Wages of Workers (Rs)	:	105-115	115-125	125-135	135-145	145-155
No. of Workers	:	8	15	25	40	62

Solution: We shall use empirical formula for the calculation of mode.

Calculation of $\overline{X}$ and M_d

Class Intervals	*Frequency*	*c. f.*	*Mid - Values*	u $\frac{X \quad 130}{10}$	*fu*
105 - 115	8	8	110	2	16
115 - 125	15	23	120	1	15
125 - 135	25	48	130	0	0
135 - 145	40	88	140	1	40
145 - 155	62	150	150	2	124
Total	150				133

Since $\dfrac{N}{2} \quad \dfrac{150}{2} = 75$, the median class is 135 - 145

$\therefore \qquad M_d = 135 + \dfrac{75 - 48}{40} \times 10 = 135 + 6.75 = 141.75$

Also $\quad \overline{X} = 130 + \dfrac{133 \times 10}{150} = 138.87$

Thus, $M_o = 3 \times 141.75 - 2 \times 138.87 = 147.51$

1.20 CHOICE OF A SUITABLE AVERAGE

The choice of a suitable average, for a given set of data, depends upon a number of considerations which can be classified into the following broad categories:

a. Considerations based on the suitability of the data for an average.
b. Considerations based on the purpose of investigation.
c. Considerations based on various merits of an average.

(a) ***Considerations based on the suitability of the data for an average:***

1. The nature of the given data may itself indicate the type of average that could be selected. For example, the calculation of mean or median is not possible if the characteristic is neither measurable nor can be arranged in certain order of its intensity. However, it is possible to

calculate mode in such cases. Suppose that the distribution of votes polled by five candidates of a particular constituency are given as below:

Name of the Candidates	:	A	B	C	D	E
No. of votes polled	:	10,000	5,000	15,000	50,000	17,000

Since the above characteristic, i.e., name of the candidate, is neither measurable nor can be arranged in the order of its intensity, it is not possible to calculate the mean and median. However, the mode of the distribution is D and hence, it can be taken as the representative of the above distribution.

2. If the characteristic is not measurable but various items of the distribution can be arranged in order of intensity of the characteristics, it is possible to locate median in addition to mode. For example, students of a class can be classified into four categories as poor, intelligent, very intelligent and most intelligent. Here the characteristic, intelligence, is not measurable. However, the data can be arranged in ascending or descending order of intelligence. It is not possible to calculate mean in this case.
3. If the characteristic is measurable but class intervals are open at one or both ends of the distribution, it is possible to calculate median and mode but not a satisfactory value of mean. However, an approximate value of mean can also be computed by making certain an assumption about the width of class (es) having open ends.
4. If the distribution is skewed, the median may represent the data more appropriately than mean and mode.
5. If various class intervals are of unequal width, mean and median can be satisfactorily calculated. However, an approximate value of mode can be calculated by making class intervals of equal width under the assumption that observations in a class are uniformly distributed. The accuracy of the computed mode will depend upon the validity of this assumption.

(b) ***Considerations based on the purpose of investigation:***

1. The choice of an appropriate measure of central tendency also depends upon the purpose of investigation. If the collected data are the figures of income of the people of a particular region and our purpose is to estimate the average income of the people of that region, computation of mean will be most appropriate. On the other hand, if it is desired to study the pattern of income distribution, the computation of median, quartiles or percentiles, etc., might be more appropriate. For example, the median will give a figure such that 50% of the people have income less than or equal to it.

   Similarly, by calculating quartiles or percentiles, it is possible to know the percentage of people having at least a given level of income or the percentage of people having income between any two limits, etc.

2. If the purpose of investigation is to determine the most common or modal size of the distribution, mode is to be computed, e.g., modal family size, modal size of garments, modal size of shoes, etc. The computation of mean and median will provide no useful interpretation of the above situations.

(c) ***Considerations based on various merits of an average:*** The presence or absence of various characteristics of an average may also affect its selection in a given situation.

1. If the requirement is that an average should be rigidly defined, mean or median can be chosen in preference to mode because mode is not rigidly defined in all the situations.
2. An average should be easy to understand and easy to interpret. This characteristic is satisfied by all the three averages.
3. It should be easy to compute. We know that all the three averages are easy to compute. It is to be noted here that, for the location of median, the data must be arranged in order of magnitude. Similarly, for the location of mode, the data should be converted into a frequency distribution. This type of exercise is not necessary for the computation of mean.
4. It should be based on all the observations. This characteristic is met only by mean and not by median or mode.
5. It should be least affected by the fluctuations of sampling. If a number of independent random samples of same size are taken from a population, the variations among means of these samples are less than the variations among their medians or modes. These variations are often termed as sampling variations.

   Therefore, preference should be given to mean when the requirement of least sampling variations is to be fulfilled. It should be noted here that if the population is highly skewed, the sampling variations in mean may be larger than the sampling variations in median.

6. It should not be unduly affected by the extreme observations. The mode is most suitable average from this point of view. Median is only slightly affected while mean is very much affected by the presence of extreme observations.
7. It should be capable of further mathematical treatment. This characteristic is satisfied only by mean and, consequently, most of the statistical theories use mean as a measure of central tendency.
8. It should not be affected by the method of grouping of observations. Very often the data are summarized by grouping observations into class intervals. The chosen average should not be much affected by the changes in size of class intervals.

It can be shown that if the same data are grouped in various ways by taking class intervals of different size, the effect of grouping on mean and median will be very small particularly when the number of observations is very large. Mode is very sensitive to the method of grouping.

9. It should represent the central tendency of the data. The main purpose of computing an average is to represent the central tendency of the given distribution and, therefore, it is desirable that it should fall in the middle of distribution. Both mean and median satisfy this requirement but in certain cases mode may be at (or near) either end of the distribution.

Practice Exercise

Q 1. What Do you understand by frequency distribution? Name the graph through which frequency distribution is drawn? **[Sep. 2018]**

Q 2. What Do you understand by frequency distribution? Why graphical distribution is helpful in frequency distribution? **[Sep. 2015]**

Q 3. Wages of 65 employees of a company for 3 days are given as follows. Find out the mean by direct method. **[Sep. 2018]**

Wages(in Rs.)	55	65	75	85	95	105	115
No. Of Employee	8	10	16	14	10	5	2

Q 4. What is average? Which average do you consider best and why? **[Sep. 2017]**

Q 5. Discuss any two of the following (a) Mean (b) Median (c) Mode **[Sep.16]**

Q 6. Find the median from the following data: 23,13,17,19,20,21,16,18,17,25 **[Sep.16]**

Q 7. Find the mean and mode of the following data: **[Sep.16]**

Marks (X)	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Student(f)	2	18	30	45	35	20	6	3

Q 8. Determine the arithmetic mean of salaries of staff members as shown in the following table: **[Sep.15]**

Employee no.	1	2	3	4	5	6
Monthly salary	12000	14500	8500	13500	13500	17500

Q 9. What are the properties of good measure of central tendency? **[Sep.14]**

Q 10. Find the median of the following data of the marks of 20 students
21,33,37,56,47,25,33,32,47,34,35,23,26,33,37,26,37,37,43,45.