INTRODUCTION OF STASTICS 26/04/2023

Statistics :

statistics is the sign which deals with the collection, analysis and interpretation of numerical data.

The main functions of statistics are discussed below:

i) Presents facts in numerical figure.

ii) To study relationship between two or more phenomenon

iii) Present complex facts in a simplified form

iv) Provide techniques for the comparision of phenomenon.

v) Helps in formation of polysis.

vi) Helps in forecaseting.

vii) Provides techniques for testing.

viii) Provides techniques for making decisions under uncertainity

1) presents facts in numerical figures

The first function of Statistics is to present a given problem in terms of numerical figures WKT the numerical presentation helps in having a better understanding of the nature of problem.

2. presents complex facts in a simplified form:
Generally problem to be investigated is presented by a large mass of numerical figures, which are very difficult to understand and remember. using various statistical methods this large data can be presented in a simplified form.

3. To study relationship between two or more phenomena:
Statistic can be used to investigate whether two or more phenomena are related.

4. Provides techiniques for the comparision of phenomena:
Many times the purpose of understanding a statistical analysis is to compare various phenomena by computing one or more measures

5. Helps in the formation of polysis:
statistical analysis is the starting point in the formation of policesis in various economic business and government activity.

6. Helps in forcasting:
The success of planning by the government or of a business depends to a large extent up on the accuracy of their

forecast. Statistics provides a scientific basis for making such forecast.

7) Provides techiques for testing the hypothesis :

A hypothesis is a 'Strateoy statement about some characteristic of a population by using some statistics techniques. It is possible to test the validity of a statement.

8) Provides techniques for making decisions under uncertainity :

Many times we faced uncertain situations, for example a person may situation where he/she has to decide wether to take his umbrella or not due to the uncertainity of rain. Answer to such problems are provided by the statistical techniques which will helps in decision making under uncertainity.

⇒ Collection of data:

Data collection is the process of gathering information about the relevant topic of research which is done by the researcher

Sources and methods of collecting data:

i) primary data.

ii) secondary data.

→ Primary data: data collected by investigator himself is called primary data.

Methods in collecting primary data:

i) direct personal interview:

In this method data is personally collected by the interviewer.

ii) Indirect oral interview:

data is collected from third parties who have information about subject of enquiry.

iii) Mailed questionaire method:

data is collected through questionaire method to the informent

iv) Telephonic interview:-

data is collected through an interview over the telephone with the interviewer

→ Secondary data: The data which have been collected by some individuals or agencies and statistically treated to draw certain conclusions i.e., data collected by someone and use by the investigator. Secondary data is already existing and not original, it is already collected from some other purpose.

27/4/2023.

Presentation of data:

Presentation of data includes classification and tabulation of data.

→ Classification of data

Classification is the process of arranging data in groups according to their resemblance. Different modes of classification are:-

i) Geographical classification.

ii) Chronological classification.

iii) Qualitative classification.

iv) Quantitative classification.

→ Geographical classification is according to place, area, region.

→ Chronological classification is according to the time (i.e., months, yearly, daily basis etc.)

→ Qualitative classification is according to the attributes of subjects or items are honesty, intelligence, qualifications etc.

→ Quantitative classification is according to the magnitude the numerical values.

Ex: Salaries, height, weight, marks etc.

Tabulation of data:

It is the process of presenting the data in rows and columns so that it can more easily being understand & can be used for further statistical analysis.

→ Objectives of tabulation of data:

i) To reduce complexity of data.

ii) To economic space.

iii) To clarify the objective of investigation:

→ Components of table:

i) Table number.

ii) Title.

iii) Caption (column headings)

iv) Stubs (row headings)

v) Body of the table.

vi) Source.

vii) Unit of measurement.

viii) Head note.

ix) Foot note.

→ General rules for tabulation:

The table should suit the size of paper usually at more rows then columns.

i) Space must be allowed for reference or any other matter which is to be included in the table.

ii) In all tables the captions and stubs should be arranged in some systamatic order. The arrangement of items basically depends upon the type of data.

iii) The point of measurement should be clearly defined and given in the table such as income in rupees or weights in pounds.

iv) The table should not be overloaded with details.

v) Percentage and ratios should be computed and shown.

vi) Abbrivation should be avoided specially in titles and headings for example: yr should not be used for years.

⇒ Difference between classification and tabulation:

classification	Tabulation
→ It is basis for tabulation.	→ It is the basis for further analysis
→ It is the basis for simplification.	→ It is the basis for presentation.
→ Data is divided into groups and sub groups on the basis of similarities & dissimilarities.	→ data is listed according to a logical sequence or related characteristics.

→ data are seperated and grouped base on a property of the data common to all values.

→ data is arranged in to columns and row base on characteristic or properties..

Graphical presentation

Graphical presentation refers to the way of presenting the data with the help of graphs.

→ Guidelines for construction of graph:

→ Title :

The heading depicting the content of the data must be provided a title for all the graphical representation.

→ Scale:

The scale selected must satisfy by all values to be plotted on the graph.

→ Index:

The index must be provided to show the scale of 'x' and 'y'- axes.

→ Source of data :

The sources of data gives the information about the data and is mentioned at the bottom of the graph.

Functions of graph:

i) The shape of the graph offers easy answers to several questions

ii) The shape of the graph gives an exact idea of the variations of the distribution trends.

iii) Graphical presentation, therefore serves as an easy technique for quick and effective comparision between two or more frequency distribution.

28/4/23.

Different types of graphs:

The graphical representation of statistical data in a chart is normally specified as statistical graph chart.

There are many kinds of graphs & charts which are used to indicate a set of data. These graphs are very helpful to recognize the statistical data. The following are some of the graphs based on statistics

1. line graph.
2. Bar graph.
3. Histograph.
4. Frequency polygon.
5. Ogive / Cummulative frequency curve.
6. Piechart

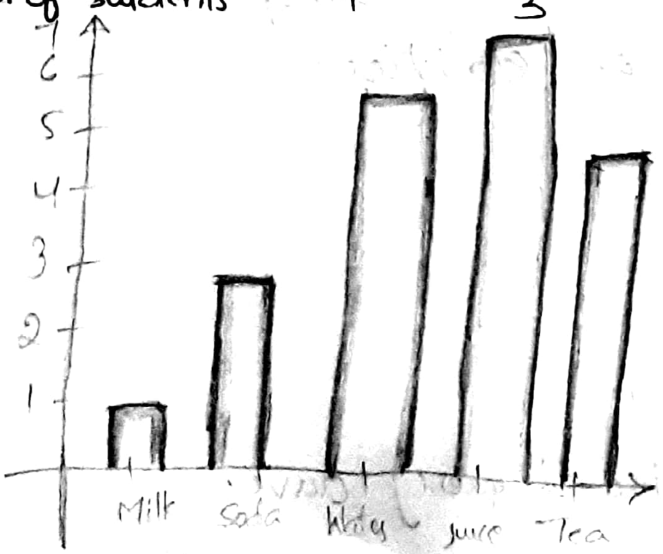
→ 3, 4, 5, 6 are called graphs of frequency distribution.

1. Line graph: A line graph is a diagram that shows a line joining several points. A line graph can be taken on xy plane where there will an independent variable and a dependent variable. Mostly the independent variable is taken on x-axis, while the dependent variable is taken on y-axis.

2. Bar graph: Bar graph is drawn on the xy plane and it have labelled horizontal or vertical bar that show different values. The size, length or column of bar represent different values. Bar graph is very useful for non-continuous data.

Ex: Draw bar graph for following data:

drink type	Milk	soda	water	juice	tea
No. of students	1	3	6.	7	5

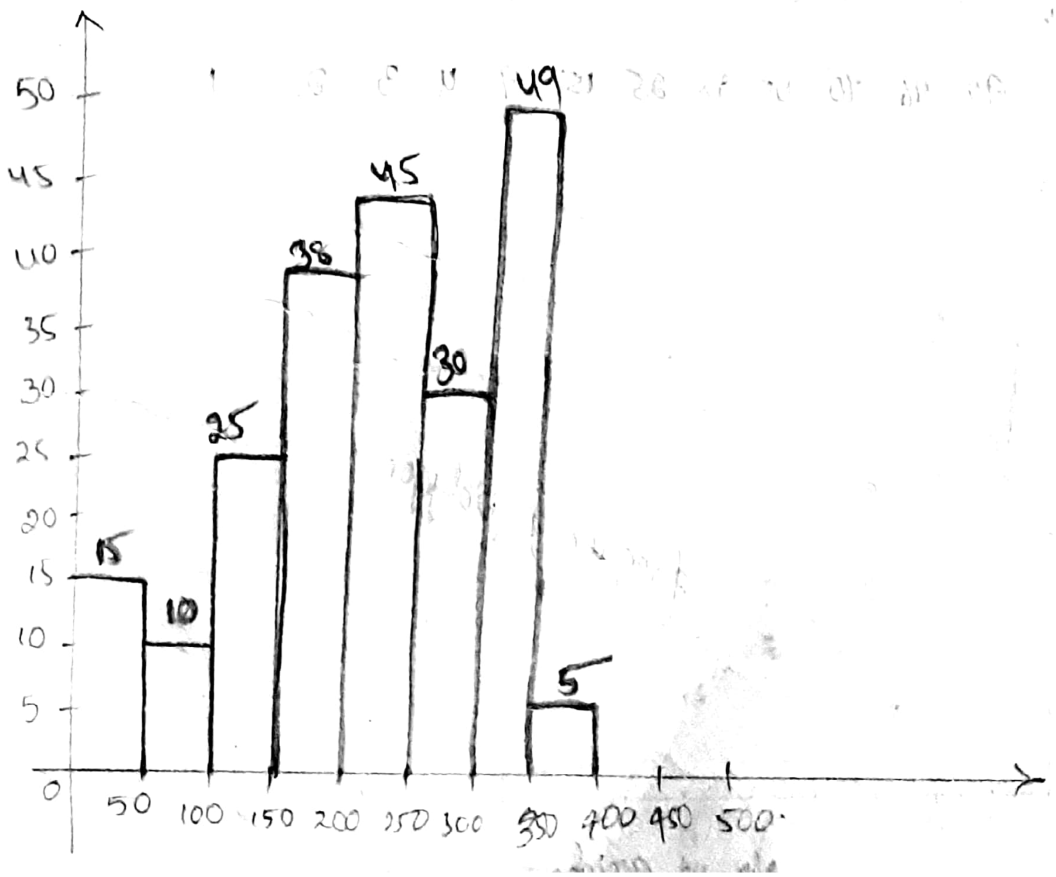


3. Histrograph: One of the most used, easily understood method for graphical representation of frequency distribution of data is called histrograph. Also known as column diagram.

→ During the construction of histrograph variables are taken on x-axis & frequencies on y-axis. If the dif. b/w class interval are same then distance b/w the ▭les on the x-axis should be same. The frequencies of each class which is equivalent to height of ▭le can be shown on y-axis.

Ex: Draw the histrograph for following data.

Marks.	0-50	50-100	100-150	150-200	200-250	250-300	300-350	350-400
No. of Students	15	10	25	38	45	30	49	5



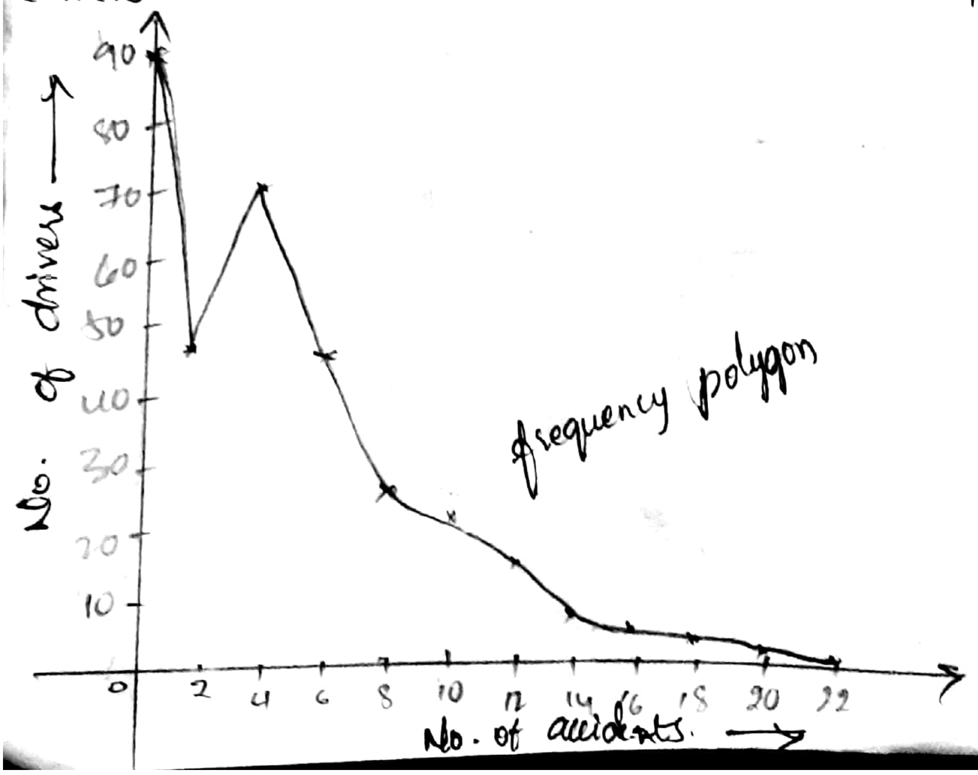
4. Frequency polygon: It is a graph of frequency distribution. There are 2 ways in which this can be constructed.

1. We can draw a histograph of given data and then join by spread lines the mid point of the upper horizontal side of each ▭le with the adj. side. The figure so formed is called frequency polygon.

2. Another method is to take the mid point of the various c.I & then plot the frequencies corresponding to each point & to join all these points by straight line. The fig obtained to be exactly same as method 1, the only difference is that we have not constructed histograph.

Ex: Draw a frequency polygon of given data.

No. of accidents	0	2	4	6	8	10	12	14	16	18	20	22
No. of drivers	90	46	70	45	28	25	15	9	4	3	2	1



01/05/2023.

Ogive (or) Cummulative frequency curve:

The Ogive curve is also known as cumulative frequency curve. There are two techniques for constructing an ogive curve

More than ogive curve: In this method we start with the lower limit of the classes and from the frequencies we subtract the frequency of each class when these frequencies are plotted we get declined curve.

Less than ogive curve: In this method we start with the upper limit of the classes and go on adding the frequencies when these frequencies are plotted we get a rising curve.

Example: Construct less than, more than ogive curves for the given data

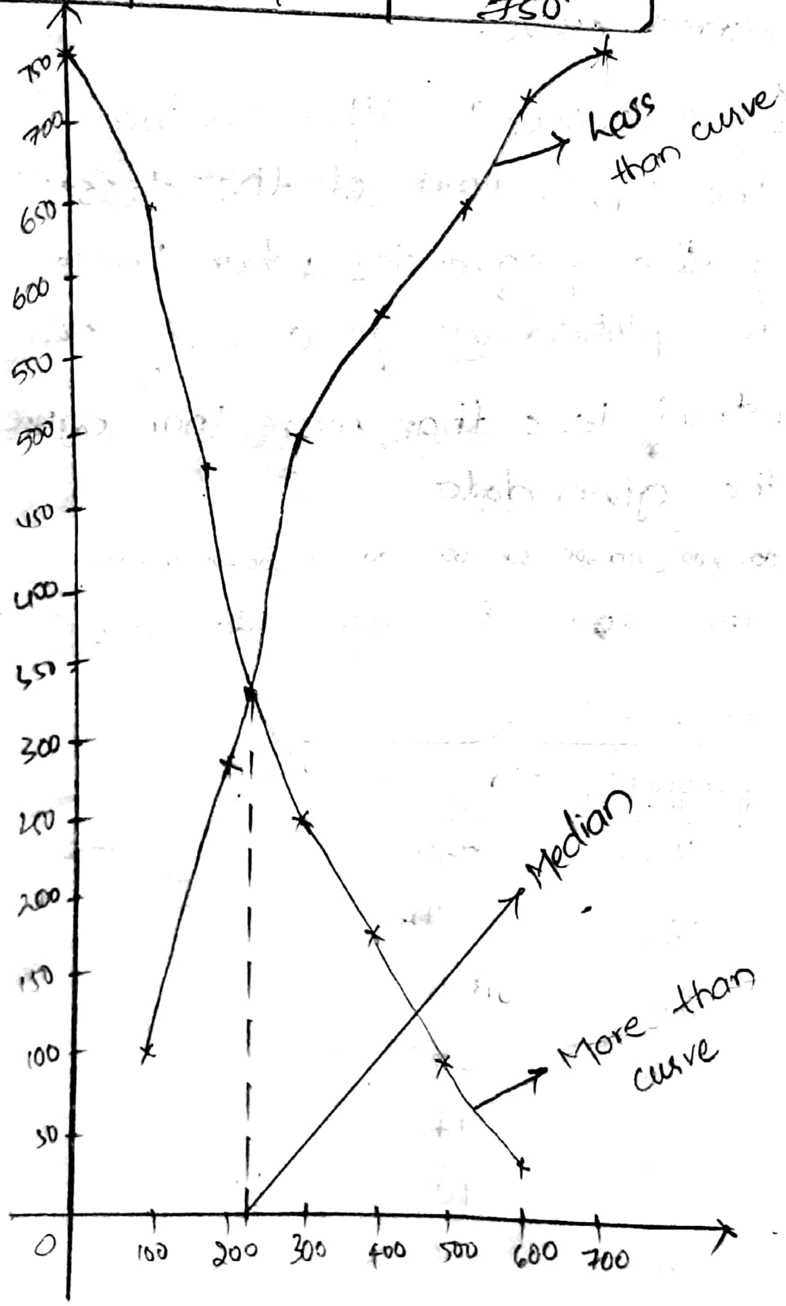
class	0-100	100-200	200-300	300-400	400-500	500-600	600-700	
frequency	100	180	220	80	70	60	40	=750=N

More than curve.

Lower Limit	frequency	More than C.F.
0	100	750
100	180	650
200	220	470
300	80	250
400	70	170
500	60	100
600	40	40.

Less than curve

Upper limit	frequency	Less than C.F
100	100	100
200	180	280
300	220	500
400	80	580
500	70	650
600	60	710
700	40	750



Example: Draw the ogive curve for the following data.

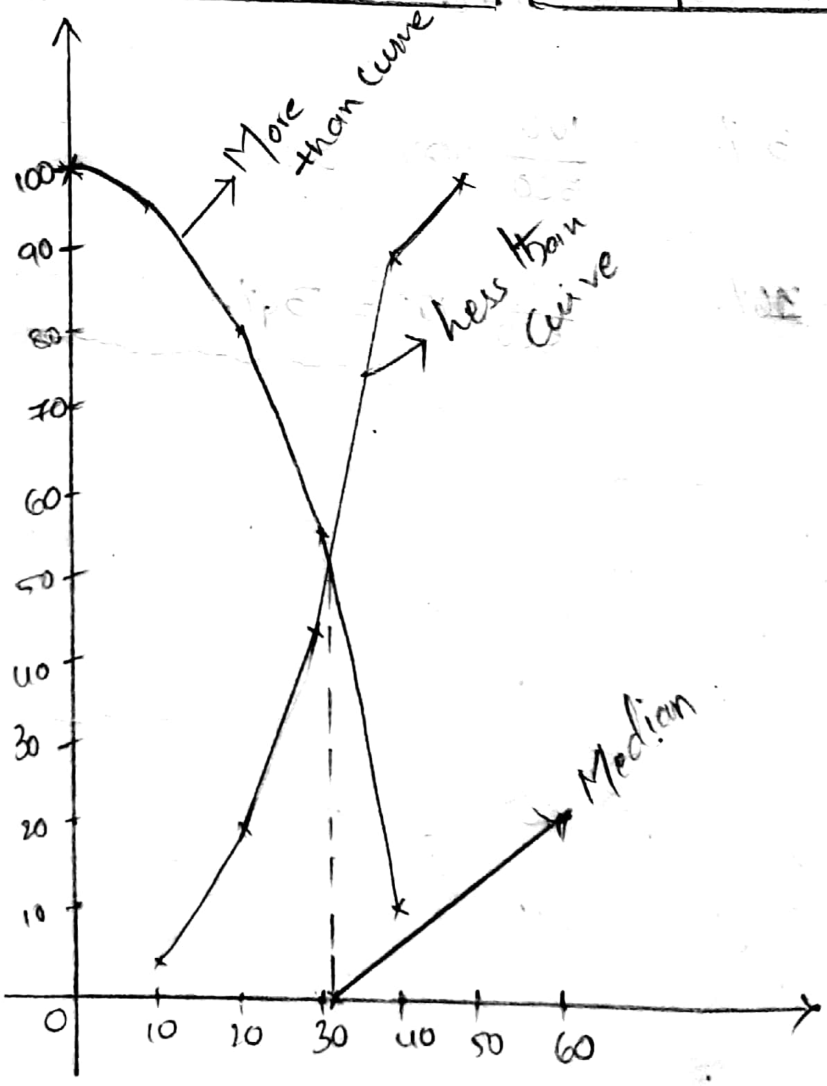
Marks	0-10	10-20	20-30	30-40	40-50	
No. of Students	4	16	24	46	10	=100 = N

More than curve

L.L	frequency	M.T.C.F
0	4	100
10	16	96
20	24	80
30	46	56
40	10	10

Less than curve.

U.L	frequency	L.T.C.F
10	4	4
20	16	20
30	24	44
40	46	90
50	10	100



Pie Chart:

A pie chart can be taken as a circular graph which is divided into different pieces each displaying the size of some related information.

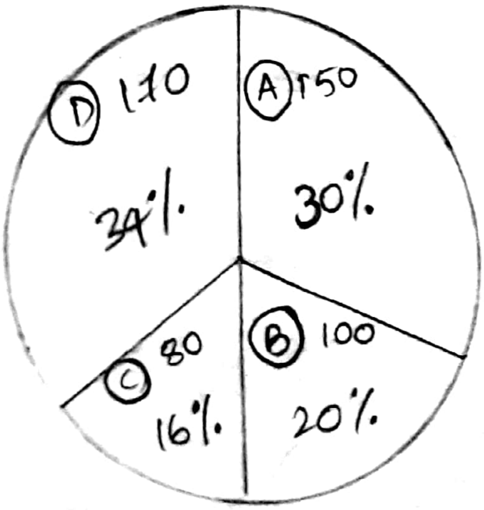
Pie charts are best use with respect to categorical data which helps in understanding what percentage each of this category constitutes

Ex: Draw a pie chart for the following data.

Cars	A	B	C	D
Sales	150	100	80	170

A) $\frac{150}{500} \times 100 = 30\%$. B) $\frac{100}{500} \times 100 = 20\%$.

C) $\frac{80}{500} \times 100 = 16\%$. D) $\frac{170}{500} \times 100 = 34\%$.



03/05/2023.

Measures of central tendency:

A measure of central tendency is an attempt to find a single value to describe the whole data.

Various types of measures of central tendency:

Various measures of central tendency are classified as follows:

1) Arithmetic mean (or) mean (or) Average.
2) Geometric mean
3) Harmonic mean
4) Median
5) Mode

Here Arithmetic mean, Geometric, Harmonic means are called mathematical averages.
Median & Mode are called positional averages

* Arithmetic Mean (A.M):

A.M of a series is the value obtain by dividing the total value of the various item by their number. In individual series A.M can be calculated as $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ [Direct method].

$$\bar{x} = A + \frac{\sum_{i=1}^{n} d_i}{n} \quad \text{[Indirect method]}$$

$d_i = x_i - A$;

A = Assumed value.

n = no. of terms

→ In discrete frequency data arithmetic mean can be calculated as

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{N} \quad \text{[Direct method]}$$

$N = \sum f_i$

(or)

$$\bar{x} = A + \frac{\sum_{i=1}^{n} f_i d_i}{N} \quad \text{[Shortcut Method]}$$

$N = \sum f_i$; $d_i = x_i - A$; A = Assumed value.

→ In continuous series arithmetic mean can be calculated as

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i m_i}{N} \quad \text{[Direct method]}$$

(or)

$$\bar{x} = A + \frac{\sum f_i d_i}{N} \times C \quad \text{[Step deviation method]}$$

A = Assumed value

m_i = Mid value of each class

$N = \sum f_i$

C = length of class

$$d_i = \frac{m_i - A}{C}$$

Merits and demerits of arithmetic mean:

→ **Merits :-**

i) It is easy to understand and easy to calculate.

ii) It takes all the values into consideration thus it is more representative.

iii) It is used in the computation of various other statistical measures.

iv) It is possible to calculate even if some of the details of the data are lacking.

v) It provides a good basis for comparision.

→ **Demerits :-**

i. It cannot be located graphically.

ii. It cannot be used in the study of qualitative phenomenon.

iii. It cannot be calculated if the data has open ended classes.

Problems :

1) Calculate A.M of following table which gives the monthly income of 10 employees in an office.

Income in (Rs) : 4,780 ; 5760 ; 6690 ; 7,750 ; 4,840 ; 4,920 ; 6,100 ; 7,810 ; 7,050 ; 6,950.

Sol:- Direct method:

$$\boxed{\bar{X} = \frac{\sum_{i=1}^{N} x_i}{n}}$$ n = no. of terms.

$n = 10$

$$\bar{x} = \frac{62650}{10} = 6265$$

Indirect method:

$$\boxed{X = A + \frac{\sum d_i}{n}}$$; $d_i = x_i - A$

$A = 4840$.

x	$d_i = x_i - A$
4780	-60
5760	920
6690	1850
7750	2910
4840 A	0
4920	80
6100	1260
7810	2970
7050	2210
6950	2110

$\sum d_i = 14250$

$$\Rightarrow X = 4840 + \frac{14250}{10}$$

$$\boxed{X = 6265}$$

04/05/2023.

From the following data of related to the distance travelled by 520 villages to buy their weekly requirement. Find the mean.

Miles travelled	No. of villages (f)	fx
2	38	76
4	104	416
6	140	840
8	78	624
10	48	480
12	42	504
14	28	392
16	24	384
18	16	288
20	2	40
	$\Sigma f_i = 520$	4044

$$\bar{x} = \frac{\sum f_i x_i}{N}, \quad N = \Sigma f_i$$

$$\bar{x} = \frac{4044}{520} = 7.77$$

shortcut method.

Miles travelled (x_i)	No. of villages f_i	$d_i = x_i - A$	$f_i d_i$
2	38	-8	-304
4	104	-6	-624
6	140	-4	-560
8	78	-2	-156
$\boxed{10} \rightarrow A$	48	0	0
12	42	2	84
14	28	4	112
16	24	6	144
18	16	8	128
20	2	10	20
	$\Sigma f_i = 520$		$\Sigma f_i d_i = -1156$

$$\boxed{\bar{x} = A + \frac{\sum f_i d_i}{N}}$$ (Indirect method)

$$\bar{x} = 10 + \frac{-1156}{520}$$

$= 7.77$

Find the avg. weights wages of workers.

wages	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of workers	170	80	50	70.	25	25	150	100.

wages	No. of workers (f_i)	Mid value (m_i)	$f_i m_i$	$d_i = \frac{m_i - A}{C}$	$f_i d_i$
0-10	170	5	850	-3	-510
10-20	80	15	1200	-2	-160
20-30	50	25	1250	-1	-50
30-40	70	$\boxed{35}$ A	2450	0	0
40-50	25	45	1125	1	25
50-60	25	55	1375	2	50
60-70	150	65	9750	3	450
70-80	100	75	7500	4	400
	$\sum f_i = 670$		$\sum f_i m_i = \boxed{25500}$		$\sum f_i d_i = 205$

$$\boxed{\mu = \frac{\sum f_i m_i}{N}} = \frac{25500}{670} = 38.0597$$

(direct method)

Step deviation method:

$$\boxed{\bar{x} = A + \left(\frac{\Sigma f_i d_i}{N} \times C\right)} \quad \left[\because d_i = \frac{m_i - A}{C}\right] \quad A = 35$$

$$\bar{x} = 35 + \frac{205}{670} \times 10$$

$$\bar{x} = 35 + 3.0597$$

$$\bar{x} = 38.0597$$

Fifth find the A.M for the following data using step deviation method.

class. Int	frequency (f_i)	(Mid val) (m_i)	~~$f_i m_i$~~	$d_i = \frac{m_i - A}{C}$	$f_i d_i$
100-200	8	150		-3	-24
200-300	16	250		-2	-32
300-400	26	350		-1	-26
400-500	20	$\boxed{450}$ (A)		0	0
500-600	12	550		1	12
600-700	10	650		2	20
700-800	8	750		3	24
	$\Sigma f_i = 100$				$\Sigma f_i d_i = -26$

$$\bar{x} = A + \left(\frac{\Sigma f_i d_i}{N} \times C\right)$$

$$= 450 + \frac{-26}{100} \times 100$$

$$= 424$$

Scanned by CamScanner

Find the mean for following data 5/5/23.

values	C.F	class	f_i	mid(m_i)	$d_i = \frac{m_i - A}{c}$	$\Sigma f_i d_i$
<10	4	0-10	4	5	-3	-12
<20	16	10-20	12	15	-2	-24
<30	40	20-30	24	25	-1	-24
<40	76	30-40	36	$\boxed{35}$ A	0	0
<50	96	40-50	20	45	1	20
<60	112	50-60	16	55	2	32
<70	120	60-70	8	65	3	24
<80	124	70-80	4	75	4	16
			$\Sigma f_i = 124$			$\Sigma f_i d_i = 32$

$$\bar{x} = A + \left(\frac{\Sigma f_i d_i}{N} \times C\right) = 35 + \frac{32}{124} \times 10$$

$$\bar{x} = 35 + 2.5806 = 37.5806.$$

Find the AM of following data. Calculate avg. marks of 50 students.

More.t.marks values	No. of Stu.	C.I	f_i	mid(m_i)	$d_i = \frac{m_i - A}{c}$	$\Sigma f_i d_i$
>0	50	0-10	4	5	-2	-8
>10	46	10-20	6	15	-1	-6
>20	40	20-30	20	$\boxed{25}$ A	0	0
>30	20	30-40	10	35	1	10
>40	10	40-50	7	45	2	14
>50	3	50-60	3	55	3	9
			$\Sigma f_i = 50$			$\Sigma f_i d_i = 19$

$$\bar{x} = A + \left(\frac{\Sigma f_i d_i}{N} \times C\right) = 25 + \left(\frac{19}{50} \times 10\right)$$

$$\bar{x} = 25 + 3.8000 = 28.8000.$$

Calculate A.M from the following data.

Marks more	No. of stu.	C.I	f_i	m_i	$d_i = \frac{m_i - A}{C}$	f_id_i
~~200~~						
>0	150	0-10	10	5	-3	-30
>10	140	10-20	40	15	-2	-80
>20	100	20-30	20	25	-1	-20
>30	80	30-40	0	35 A	0	0
>40	80	40-50	10	45	1	10
>50	70	50-60	40	55	2	80
>60	30	60-70	16	65	3	48
>70	14	70-80	14	75	4	56
			$\Sigma f_i = 150$			$\Sigma f_id_i = 64$

$$\bar{x} = A + \left(\frac{\Sigma f_id_i}{N} \times C\right) = 35 + \left(\frac{64}{150} \times 10\right)$$

$$\bar{x} = 35 + 4.2667 = 39.2667$$

Calculate A.M from the following data

marks	freq	C.I	f_i	m_i	$d_i = \frac{m_i - A}{C}$	f_id_i
<10	4	0-10	4	5	-2	-8
<20	20	10-20	16	15	-1	-16
<30	40	20-30	20	$\boxed{25}$ A	0	0
<40	40	30-40	0	35	1	0
<50	50	40-50	10	45	2	20
			$\Sigma f_i = 50$			$\Sigma f_id_i = -4$

$$\bar{x} = A + \left(\frac{\Sigma f_id_i}{N} \times C\right)$$

$$\bar{x} = 25 + \left(\frac{-4}{50} \times 10\right)$$

$$\bar{x} = 25 - 0.8000 = 24.2000$$

08/05/28.

Find the missing frequencies from the following data when the average ($\bar{x} = 16.82$)

Classes	f_i	m	$d_i = \frac{m_i - A}{c}$	$f_i d_i$
0-5	10	2.5	-3	-30
5-10	12	7.5	-2	-24
10-15	16	12.5	-1	-16
15-20	x	$\boxed{17.5}$ A	0	0
20-25	14	22.5	1	14
25-30	10	27.5	2	20
30-35	8	32.5	3	24
	$\Sigma f_i = 70+x$			$\Sigma f_i d_i = -12$

$$\bar{x} = A + \left(\frac{\Sigma f_i d_i}{N} \times C\right) = 17.5 + \left(\frac{-12}{70+x} \times 5\right)$$

$$16.82 = 17.5 + \frac{-60}{70+x}$$

$$\cancel{-}0.68 = \cancel{-}\frac{60}{70+x}$$

$$(0.68)(70+x) = 60$$

$$70+x = \frac{60}{0.68}$$

$$70+x = 88.2353$$

$$\boxed{x = 18.2353}$$

Find the missing frequency when $\bar{x} = 28$.

Class	f_i	x_i	$d_i = \frac{x_i - A}{C}$	$f_i d_i$
0-10	12	5	-3	-36
10-20	18	15	-2	-36
20-30	27	25	-1	-27
30-40	?	$\boxed{35}$ A	0	0
40-50	17	45	1	17
50-60	6	55	2	12
	$\Sigma f_i = 80 + y$			$\Sigma f_i d_i = -70$

$$\bar{X} = A + \left(\frac{\Sigma f_i d_i}{N} \times C \right)$$

$$28 = 35 + \left(\frac{-70}{80 + y} \times 10 \right)$$

$$-7 = \frac{-700}{80 + y}$$

$$80 + y = \frac{\cancel{700}\ 100}{\cancel{7}}$$

$$80 + y = 100$$

$$\boxed{y = 20}$$

If the mean of the following data is 65.6 find the missing frequencies.

Scanned by CamScanner

class	f_i	m_i	$f_i m_i$
10-30	5	20	100
30-50	2	40	80
50-70	x	60	$60x$
70-90	20	80	1600
90-110	y	100	$100y$
110-130	2	120	240
Total	50		$\boxed{\Sigma f_i m_i = 2020 + 60x + 100y}$

$$\left(\text{Mean} = \frac{\text{Sum of terms}}{\text{No. of terms}}\right)$$

$\Sigma f_i = 29 + x + y = 50 \Rightarrow x + y = 21$ — (1)

$$\boxed{\bar{x} = \frac{\Sigma f_i m_i}{\Sigma f_i}}$$

$$65.6 = \frac{2020 + 60x + 100y}{50}$$

$$3280 = 2020 + 60x + 100y$$

$$1260 = 60x + 100y$$

$$3x + 5y = 63 \quad \text{— (2)}$$

Solve (1) & (2)

$x + y = 21$ Multiply eqn. (1) $\times 3$.

$3x + 5y = 63$

$$\Rightarrow \begin{array}{r} \cancel{3x} + 3y = 63 \\ \cancel{3x} + 5y = 63 \\ \hline -2y = 0 \end{array}$$

$\boxed{y = 0} \Rightarrow x + 0 = 21$

$\therefore \boxed{x = 21}$

Unequal class interval: 5/6/23.

Find the A.M for the following data:

C.I	f
0-2	2
2-5	4
5-8	7
8-10	8
10-14	10
14-15	13
15-17	3
17-20	2
20-25	1

C.I	f_i	m_i	$f_i m_i$
0-5	6	2.5	15
5-10	15	7.5	112.5
10-15	23	12.5	287.5
15-20	5	17.5	87.5
20-25	1	22.5	22.5
			525

$\Sigma f_i = 50$

$\Sigma f_i m_i = 525$

$$\bar{x} = \frac{\Sigma f_i m_i}{\Sigma f_i} = \frac{525}{50} = 10.5$$

A.M = 10.5.

Geometric mean:

Geometric mean is defined as n^{th} root of the product of 'n' values.

* Calculation of G.M for Individual series:

$$G.M = \text{Antilog}\left[\frac{\Sigma \log x_i}{N}\right]$$

* Calculation of geometric mean for discrete data:

$$G.M = \text{antilog}\left[\frac{\Sigma f_i \log x_i}{N}\right] \qquad N = \Sigma f_i$$

* calculation of geometric mean for continuous data:

$$G.M = \text{Antilog}\left[\frac{\Sigma f_i \log m_i}{N}\right] \qquad N = \Sigma f_i.$$

Merits and demerits of geometric mean:

Merits:

1) It is based on all the observations.
2) It is capable of mathematical treatment
3) It gives more weight to smaller observations & vice versa.

Demerits:

1. It is not very easy to calculate & hence not very popular.
2. It cannot be calculated if anyone of the observation is zero or negative.

Q: Find the G.M for the following data.

x. 1 7 29 92 115 375

x	$\log x$
1	0
7	0.84
29	1.46
92	1.96
115	2.06
375	2.57
$\Sigma = n$	$8.89 \; \Sigma \log x_i$

$$G.M = \text{antilog}\left[\frac{\Sigma \log x_i}{n}\right]$$

$$G.M = \text{antilog}\left[\frac{8.89}{6}\right]$$

$$G.M = \text{antilog}\,(1.48)$$

$$G.M = 10^{1.48}$$

$$\underline{G.M = 30.19}$$

Q. Find out G.M from the following data.

x	f	$\log x_i$	$f_i \log x_i$
10	12	1	12
20	15	1.30	19.50
30	25	1.47	36.75
40	10	1.60	16.
50	6	1.69	10.14
60	2	1.77	3.54

$$G.M = \text{antilog}\left[\frac{\Sigma f_i \log x_i}{N}\right]$$

$\Sigma f_i = 70$

$\Sigma f_i \log x_i = 97.93$

$$G.M = \text{antilog}\left[\frac{97.93}{70}\right]$$

$$G.M = \text{antilog}\,(1.399)$$

$$G.M = 10^{1.399}$$

$$G.M = 25.06$$

Scanned by CamScanner

Q: Calculate G.M for the following data.

x_i	f_i	$\log x_i$	$f_i \cdot \log x_i$
5	13	0.6990	9.0870
10	18	1	18
15	50	1.1761	58.8050
20	40	1.3010	52.04
25	10	1.3979	13.979
30	6	1.4771	8.8626.

$\Sigma f_i = 137.$

$\Sigma f_i \log x_i = 160.7736$

$$G.M = \text{antilog}\left[\frac{\Sigma f_i \log x_i}{N}\right]$$

$$G.M = \text{antilog}\left[\frac{160.7736}{137}\right]$$

$$G.M = \text{antilog}(1.1735)$$

$$G.M = 10^{1.1735}$$

$$G.M = 14.9108$$

Q

C.I	10–20	20–30	30–40	40–50	50–60	60–70
f.	8	12	20	10	7	3.

C.I	f_i	m_i	$\log m_i$	$f_i \log m_i$
10-20	8	15	1.1760	9.4080
20-30	12	25	1.3979	16.7748
30-40	20	35	1.5441	30.8820
40-50	10	45	1.6532	16.5320
50-60	7	55	1.7404	12.1828
60-70	3	65	1.8129	5.4387

$\Sigma f_i \log m_i = 91.2183$

$G.M = \text{Antilog} \frac{\Sigma f_i \log m_i}{N}$

$= \text{Antilog} \frac{91.2183}{60}$

$= \text{Antilog}(1.5203) = 10^{1.5203}$

$= 33.1360.$

7/6/23

Harmonic mean: The Reciprocal of the A.M of the reciprocal of the individual Observations is known as Harmonic mean.

Harmonic mean for individual series.

$$H.M = \frac{n}{\Sigma 1/x}$$

Harmonic mean for discrete data.

$$H.M = \frac{N}{\Sigma f/x} \quad ; N = \Sigma f_i$$

Harmonic mean for continuous series.

$$H.M = \frac{N}{\Sigma f/m} \quad ; N = \Sigma f_i$$

Merits & demerits of Harmonic mean:

Merits:

1) It is based on all the observations.

2) It is capable of further mathematical treatment.

3) It is suitable in computing avg. rate under

certain conditions.

Demerits:

1) It is not easy to compute and is difficult to understand.

2) It cannot be calculated in one or more observations are equal to zero.

1. Calculate H.M for the following data.

15, 17, 20, 1, 5, 2, 6, 10.

x	$1/x$
15	$1/15 = 0.0667$
17	0.0588
20	0.0500
1	1
5	0.2000
2	0.5000
6	0.1667
10	0.1
	$\Sigma 1/x = 2.1422$

$$H.M = \frac{n}{\Sigma 1/x}$$

$$= \frac{8}{2.1422}$$

$$= 3.7345$$

2. Calculate H.M for following data.

x	$1/x$
6	0.1662
9	0.1111
18	0.0556
20	0.0500
5	0.2000
6	0.1662
9	0.1111
8	0.1250
10	0.1000
11	0.0909

$\Sigma 1/x = 1.771$

$$H.M = \frac{n}{\Sigma 1/x}$$

$$= \frac{10}{1.771}$$

$$H.M = 8.4955$$

3. Calculate the H.M for the following.

i)

x	10	20	25	30	35	40
f	2	5	6	7	8	10

x	f	f/x
10	2	0.2000
20	5	0.2500
25	6	0.2400
30	7	0.2333
35	8	0.2286
40	10	0.2500

$N = 38 \quad \Sigma f/x = 1.4019$

$$H.M = \frac{N}{\Sigma f/x}$$

$$= \frac{38}{1.4019}$$

$$= 27.1061$$

ii)

C.I	0-10	10-20	20-30	30-40	40-50	50-60	60-70
f	5	20	15	30	35	40	8

x	f	m	f/m
0-10	5	5	1
10-20	20	15	1.3333
20-30	15	25	0.6000
30-40	30	35	0.8571
40-50	35	45	0.7778
50-60	40	55	0.7273
60-70	8	65	0.1231

$N = 153 \quad \Sigma f/m = 5.4186$

$$H.M = \frac{N}{\Sigma (f/m)}$$

$$= \frac{153}{5.4186}$$

$$= 28.2361$$

iii)

x	25	36	48	50	100
$1/x$	0.04000	0.0278	0.0208	0.0200	0.0100

x	120	105
$1/x$	0.0083	0.0095

$$H.M = \frac{n}{\Sigma 1/x} = \frac{7}{0.1364} = 51.3196$$

Scanned by CamScanner

iv)

x	10	11	12	13	14
f	5	8	10	9	6

x	f	f/x
10	5	0.5000
11	8	0.7273
12	10	0.8333
13	9	0.6923
14	6	0.4286

$N = 38$ $\quad \Sigma f/x = 3.1815$

$$H.M = \frac{N}{\Sigma f/x}$$

$$= 38/3.1815$$

$$= 11.9441$$

v)

C.I	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
f	5	8	11	21	35	30	22	18

x	f	m	f/m.
0-10	5	5	1.00
10-20	8	15	0.5333
20-30	11	25	0.4400
30-40	21	35	0.6000
40-50	35	45	0.7778
50-60	30	55	0.5455
60-70	22	65	0.3385
70-80	18	75	0.2400

$N = 150$; $\Sigma f/m = 4.4751$

$$H.M = \frac{N}{\Sigma f/m} = \frac{150}{4.4751}$$

$H.M = 33.5188$.

Median: Median is the value which divides a series into two equal parts. In case of individual series median can be calculated as follows:

i) Arrange the data in ascending / descending order

ii) In case of odd no. of values the middle value becomes median.

iii) In case of even no.s of values the avg. of two middle values becomes median.

Median in case of discrete series:

1) Find the C.F

2) Find $\frac{N}{2}^{th}$ term where $N = \Sigma f_i$

3) Now look at the C.F and find the total which is either equal to size of $N/2$ or next higher to that and determine the variable corresponding to it which will give the value of median.

Median for continuous series:

1) Calculate C.F

2) Find $N/2$ term where N = Total frequency

3) In C.F, find the value of $N/2$ equal to or just greater than then the median can be calculated as.

$$\text{Median} = l + \left[\frac{\frac{N}{2} - m}{f}\right] \times c.$$

where,

l = lower limit of median class.

m = C.F of preceding median class.

f = frequency of median class.

c = C.I of median class.

1: Find median of following data.

22, 26, 14, 30, 18, 11, 35, 41, 12, 32.

Data in ascending order:

11, 12, 14, 18, 22, 26, 30, 32, 35, 41

$$\text{Median} = \frac{22+26}{2} = 24.$$

2. Find the median of following data.

Income (x)	No. of persons (fi)	C.F
5000	15	15
5500	20	35
5800	5	40
6000	5	45
6800	15	60
7000	10	70
7500	20	90
	$\Sigma f_i = N = 90$	

$$\frac{N}{2} = \frac{90}{2} = 45$$

$\therefore$ Median = 6000.

3. Calculate median of the following data.

C.I	f.	c.f
0-10	100	100
10-20	105	205
20-30	112	317
30-40	90	407
40-50	55	462
50-60	60	522
60-70	150	672

l = 30

$N/2 = \frac{672}{2} = 336.$

$$\text{Median} = l + \left[\frac{N/2 - m}{f}\right] \times c$$

$$= 30 + \left[\frac{336-317}{90}\right] \times 10$$

$$= 32.1111$$

4. Find median for following data.

i) Data in ascending order:

60, 61, 61, 62, 63, 63, 63, 64, 64, 64, 64, 65, 65, 66

$$\text{Median} = \frac{63+64}{2} = 63.5$$

ii)

Size (x)	frequency	C.F
10	7	7
20	12	19
30	17	36
40	29	65
50	31	96
60	21	117
70	18	135
80	5	140

$\frac{N}{2} = \frac{140}{2} = 70$

Median = 50.

iii)

Avg deposit balance	No. of deposits	C.F
0-100	26	26
100-200	68	94
200-300	145	239 m
300-400	242 f	481
400-500	188	669
500-600	65	734
600-700	16	750

$$\frac{N}{2} = \frac{750}{2} = 375$$

$$\text{Median} = l + \left[\frac{N/2 - m}{f}\right] \times c$$

$$= 300 + \left(\frac{375 - 239}{242}\right) \times 100$$

$$= 300 + 56.1983 = 356.1983$$

Merits and demerits of median:

Merits:

1) It is easy to calculate & simple to understand particularly in a series of individual observations and discrete series.

2) It is capable of further algebraic treatment.

3) It can be determined graphically.

4) Median can be calculated in case of open end classes.

Demerits:

1) It is not based on all the observations.

2) For calculations it is necessary to arrange the data whereas, in other averages we do not need any such arrangements.

3) Median is effected more by simplying fluctuations than the mean.

12/06/23.

Mode:

Mode is the value of the variable which occurs maximum number of times in a distribution. In individual series mode is the value which repeating Maximum number of times.

→ In discrete series mode is the value of x which is corresponding to highest frequency.

→ In continuous series mode can be calculated by using the formula:

$$\text{Mode} = l + \left[\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right] \times c$$

where, f_1 is the highest frequency and the class corresponding to this f_1 is the model class.

(l) L = The limit (lower limit) of the model class

f_0 = preceeding value of f_1

f_2 = Succeeding value of f_1

c = class interval of the model class.

If the highest frequency is repeated more than one time and denominator value of the formula in the continuous series will be zero then, we can use the

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

Merits and demerits of the Mode:

Merits:

1) Like mean or median it is not effected by extreme observations, it can be calculated even if these extreme values are not known.

2) It can be determined even if distribution has open end classes.

3) It can be located even when the class intervals are unequal.

4) It is the value around with there is more concentration of observation.

Demerits:

1) It is not base on all the observations.

2) It is not capable of futher mathematical treatment.

3) It is not easy to calculate unless the no. of obsevations are sufficiently large.

1) Find the mode of the following data.

10, 12, 13, 10, 18, 16, 15, 10, 11, 17, 10, 16, 11, 10, 12, 18, 19, 20, 15, 9.

Among all the numbers 10 is repeated at maximum at 5 times. Hence, Mode = 10.

2) Calculate Mode of the following data.

Size of Garments (x)	No. of persons. (f)
27	10
28	20
29	40
30 ←	65 ←
31	50
32	15

$\boxed{\text{Mode} = 30.}$

From the data the highest frequency is 65 and the corresponding value of the size of Garment to this 65 is 30. So,

mode = 30.

3) Find the mode for the following data.

classes	frequency	
0-10	4	
10-20	13	
20-30	21	f_0
30-40 ←	44	f_1
40-50	33	f_2
50-60	22	
60-70	7	

$l = 30,\ f_1 = 44,\ f_0 = 21,\ f_2 = 33,\ C = 10.$

$$\text{Mode} = l + \left[\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right] \times C$$

$$= 30 + \left[\frac{44 - 21}{2(44) - 21 - 33}\right] \times 10$$

$$= 30 + 0.6764 \times 10 = 30 + 6.764$$

$$= 36.764.$$

4. Calculate mode for the following data.

class	frequency (f)	m	fimi	CF
0-100	10	50	500	10
100-200	22	150	3300	32
200-300	55	250	13750	87
300-400	10 f_0	350	3500	97
400-500	← 65. (f_1)	450	29250	162
500-600	52 f_2	550	28600	214
600-700	65	650	42250	279
700-800	11	750	8250	290
	N=290.		129400	

$N/2 = 145$

In this problem the highest frequency 65 is repeated 2 times. Hence, the formula to calculate mode is

$$\boxed{\text{mode} = 3\,\text{median} - 2\,\text{Mean.}}$$

$$\text{Mean} = \frac{\Sigma f_i m_i}{N} = \frac{129400}{290}$$

$\frac{N}{2} = 145.$

$$\text{Median} = l + \left[\frac{\frac{N}{2} - m}{f}\right] \times C$$

$$= 400 + \left[\frac{145-97}{65}\right] \times 100$$

$$= 400 + 0.7384 \times 100$$

$$= 400 + 73.84$$

$$= 473.84$$

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

$$= 3(473.84) - 2(446.206)$$

$$= 1421.52 - 892.412$$

$$= 529.108$$

5) Find the mode for the following data

Income	No. of persons	
0-100	5	
100-200	7	
200-300	12	f_0
300-400	← 18	f_1
400-500	16	f_2
500-600	10	
600-700	5	

$l = 300$

$f_0 = 12$

$f_1 = 18$

$f_2 = 16$

$C = 100$

$$\text{Mode} = l + \left[\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right] \times C.$$

$$= 300 + \left[\frac{18-12}{2\times18 - 12 - 16}\right] \times C$$

$$= 300 + [0.75] \times 100.$$

$$= 375$$

c)

Size	frequency (f)	m	fimi	C.F.
0-5	9	2.5	22.5	9
5-10	12	7.5	90	21
10-15	15	12.5	187.5	36
15-20	16	17.5	280	52 m
20-25	17	22.5	382.5	69
25-30	17	27.5	467.5	86
30-35	10	32.5	325	96
35-40	13	37.5	487.5	109
	N = 109 $\frac{N}{2} = 54.5$		2242.5	

In this problem the highest frequency is repeated twice. So,

$$\boxed{\text{mode} = 3\,\text{median} - 2\,\text{mean}}$$

$$\text{Mean} = \frac{\Sigma f_i m_i}{N}$$

$$= \frac{2242.5}{109} = 20.573$$

$$\text{Median} = 20 + \left[\frac{54.5 - 52}{17}\right] \times 5$$

$$= 20 + 0.7352 = 20.7352$$

$$\text{Mode} = 3(20.7352) - 2(20.573)$$

$$= 62.205 - 41.470$$

$$= 20.735$$