

Statistics

Statistics is a science which deal with population, by population, we mean, a collection of observation relating to any variable. Population may be finite or infinite. So entire population is studied through a sample of population.

Sample, is a portion of the population which is selected to represent the population very closely.

1. Statistical Processes

To treatment of the statistical processes, we follow the steps below: -

- a. Collection of Data
- b. Presentation of Data
- c. Studying of Data mathmatically, by applying statistical methods.
- d. Interpretation and Conclusion of Data.

Presentation of Data :

The data when collected, should be presented in an intelligible form. Usually the data is large in number.

A frequency table is formed with first column giving variates and second column giving the frequency. Frequency is the number of times each variate is repeated.

Ex 3, 7, 4, 0, 2, 9, 7, 5, 6, 5, 8, 7, 4, 3, 4
5, 0, 1, 1, 3, 4, 7, 6, 8, 7.

Frequency Table

variate	frequency
0	2
1	2
2	1
3	3
4	4
5	3
6	2
7	5
8	2
9	1

If the population is very large, the variates are grouped in classes, usually of equal intervals. This is done to obtain the :

range = max. value - min. value among data

and number of class is $m = 1 + 3.3 \log N$

where, N number of individuals in population

Ex Represent the following 90 observations in a frequency table with suitable class interval

13 28 42 24 13 35 36 30 25
 30 36 30 25 37 30 43 24 25
 18 39 21 37 40 31 40 33 31
 45 30 47 17 49 29 34 23 30
 26 42 34 43 35 22 28 32 26
 33 21 27 18 38 28 29 30 31
 22 27 30 12 24 46 36 30 31
 14 25 16 17 53 19 42 16 17
 27 42 26 33 21 29 27 20 41
 27 36 41 30 18 26 34 33 29

Solution

The smallest number = 12

The greatest number = 53

The range $r = 53 - 12 = 41$

number of classes $m = 1 + 3.3 \log N = 1 + 3.3 \log 90 = 7$

the length of each class = $\frac{r}{m} = \frac{41}{7} = 5.8 \approx 6$

Class	Tally marks	Frequency	Class mid-value
12 - 17		9	14.5
18 - 23		11	20.5
24 - 29		23	26.5
30 - 35		24	32.5
36 - 41		12	38.5
42 - 47		9	44.5
48 - 53		2	50.5

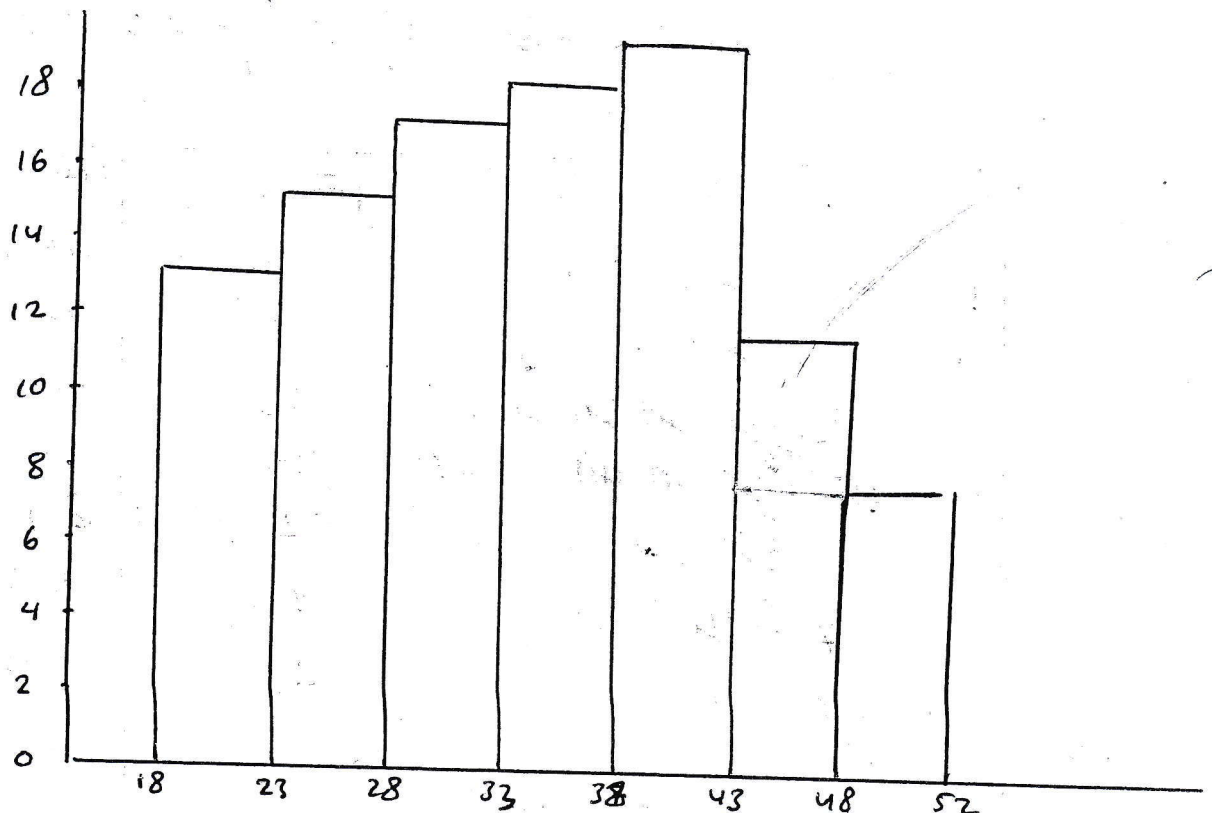
Graphical Representation

Visual aids, like diagrams, charts and graphs, have concrete form and are easily assimilated and remember.

1. Bar chart : consist of bars of equal width and different lengths, and also called Histogram.

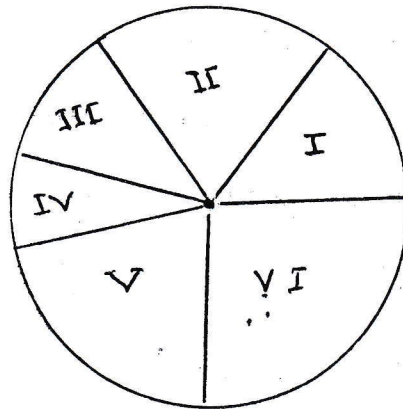
Ex

class	Frequency
18 - 22	13
23 - 27	15
28 - 32	17
33 - 37	18
38 - 42	19
43 - 47	11
48 - 52	7



2. Pie Diagram : Is a circle divided into sectors proportional in area to different item.

Items	plane	angle at center	plane %
I	372	57°	16
II	395	60°	17
III	266	41°	11
IV	179	27°	7.5
V	556	85°	23.5
VI	588	90°	25
Total	2356	360°	100



ملاحظه
نبد أن نصف القطر الافقيه باتجاه اليمين
ونستمر بجانب عقارب الساعة

3. Frequency Graph :

Dot diagram is formed when variates are noted along x-axis and dots are placed parallel to y-axis against each variate equal to the corresponding frequency in number.

Frequency polygon is got by joining the top ends of lines by means of straight lines.

Line diagram is obtained, if lines are drawn parallel to y-axis against each variate equal in length to the corresponding frequency.

Frequency Curve, when free hand smooth curve is drawn through these top ends.

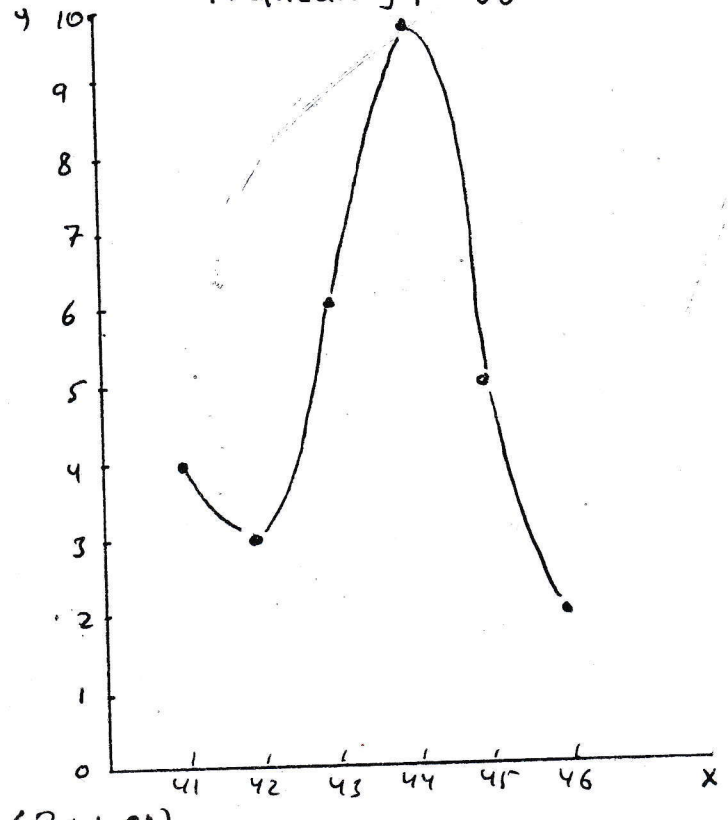
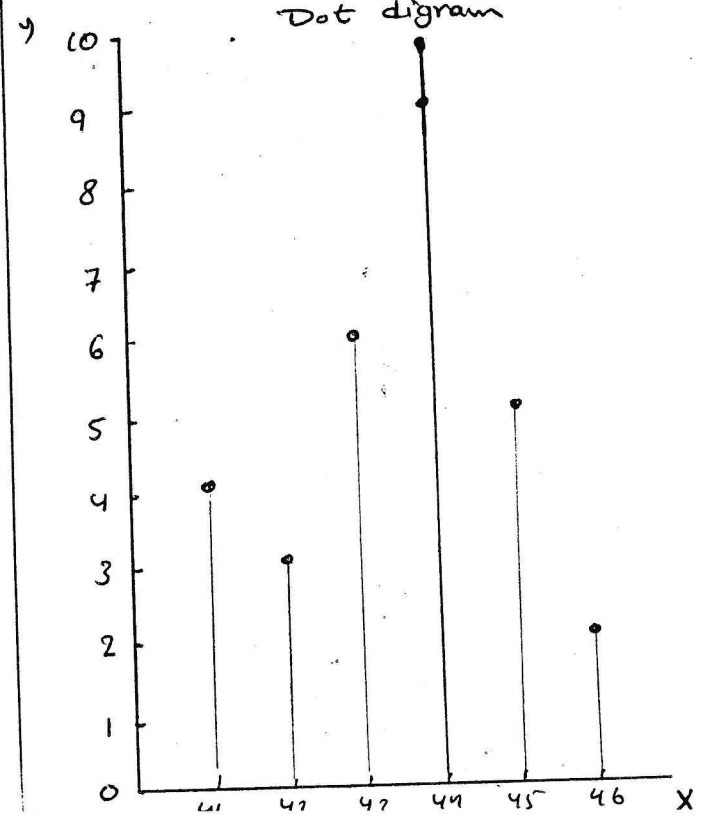
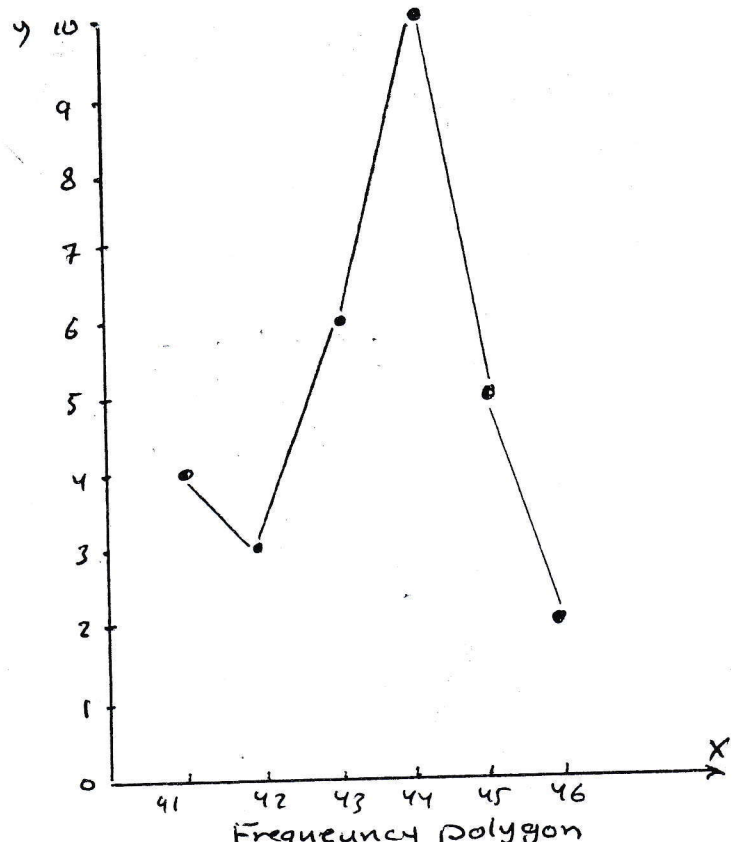
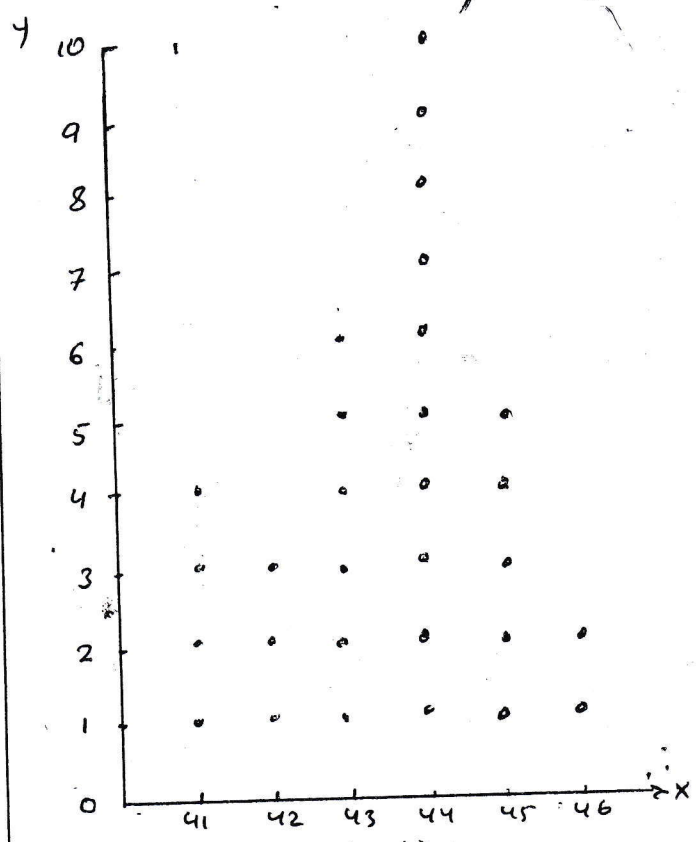
Histogram is obtained by constructing rectangles with their bases on class intervals marked on x-axis and with their areas proportional to the corresponding frequencies.

EX Draw frequency graphs to represent the following sample giving tensile strength of sheet steel in $\frac{\text{Kg}}{\text{mm}^2}$

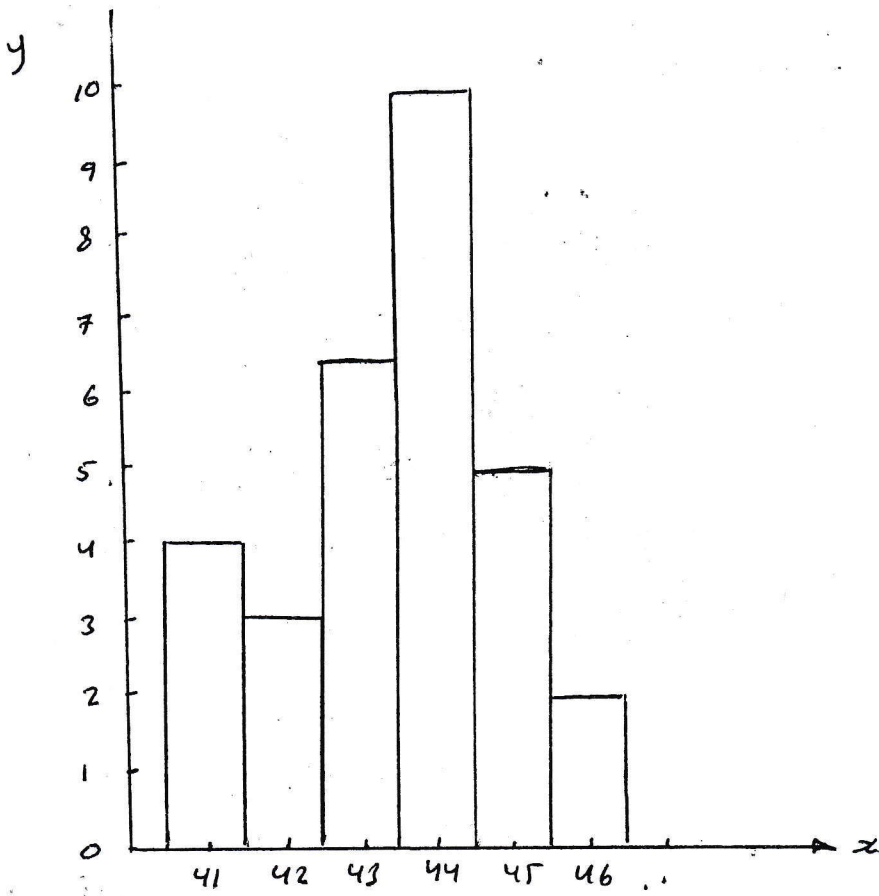
44 43 41 41 44 44 43 44 42 45
43 43 44 45 46 42 45 41 44 44
43 44 46 41 43 45 45 42 44 44

sol The frequency table is :

variate	Tally marks	frequency
41		4
42		3
43		6
44		10
45		5
46		2



(349)



Histogram

2. Measures of Location

The important parameters measure the characteristics of location and dispersion.

Averages are measure of location or central tendency. They indicate the position of centers of data. The average may be or may not be one of the values of the variate of distribution. There are five measures of central tendency :

- (i) Arithmetic mean, or Mean
- (ii) Geometric Mean
- (iii) Harmonic Mean
- (iv) Median
- (v) Mode المنوال

(i) Arithmetic mean

Also called "Mean". It is the sum of a set of observations divided by the number of observation.

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

where :

\bar{X} : Mean

X_i : observation ($i = 1, 2, 3, \dots, n$)

n : number of observation

If f_i is frequency of variate x_i , then

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

where $N = \sum f_i =$ total number of observation.

Ex Table below represent loaded applied on block
Find the Mean of loaded.

class (wt.), x	class center (x)	f given	$f x$
45 - 49	47	4	188
50 - 54	52	10	520
55 - 59	57	11	627
60 - 64	62	6	372
65 - 69	67	6	402
70 - 74	72	2	144
75 - 79	77	1	77
		$\Sigma f = 40$	$\Sigma fx = 2330$

$$\therefore \bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{2330}{40} = 58.25$$

(6)

Note If \bar{X}_1, \bar{X}_2 are the means of n_1 and n_2 observation, then the mean of the combined group of $n_1 + n_2$ is given by :

$$\bar{X} = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2}{n_1 + n_2}$$

The same can be extended to k numbers of samples with sizes, n_1, n_2, \dots, n_k and with mean $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_k$, so,

$$\bar{X} = \frac{\sum_{i=1}^k n_i \bar{X}_i}{\sum_{i=1}^k n_i}$$

Ex For a group of 60 student the mean marks in a test is 55.5. For a group of 55 student the mean is 62. Find the mean for the combined group?

sol.

$$\bar{X} = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2}{n_1 + n_2}$$

$$\bar{X}_1 = 55.5 \quad n_1 = 60$$

$$\bar{X}_2 = 62 \quad n_2 = 55$$

$$\therefore \bar{X} = \frac{(60 \times 55.5) + (55 \times 62)}{60 + 55} = 58.6$$

(iii) Geometric Mean

Is the N^{th} root of the product of N given items.

Then,

$$g = (x_1^{f_1} \cdot x_2^{f_2} \cdot x_3^{f_3} \cdot \dots \cdot x_n^{f_n})^{\frac{1}{N}}$$

where: g = geometric mean

f_1, f_2, \dots, f_n : frequencies

$$N = f_1 + f_2 + f_3 + \dots + f_n$$

$$\therefore \log g = \frac{f_1 \log x_1 + f_2 \log x_2 + \dots + f_n \log x_n}{N}$$

$$\therefore \log g = \frac{\sum f \log x}{\sum f}$$

For no frequency, $f = 1$

$$\therefore \log g = \frac{\sum \log x}{N}$$

Geometric mean, used in growth problems.

Ex The population (in millions) in the first eight censuses ^(in millions) was as follows: 3.9, 5.3, 7.2, 9.6, 12.9, 17.1, 23.2, 31.4

sol.

$$\log g = \frac{\log 3.9 + \log 5.3 + \log 7.2 + \log 9.6 + \log 12.9 + \log 17.1 + \log 23.2 + \log 31.4}{8}$$
$$= 1.0451$$

$\therefore g = 11.1$, the arithmetic mean $\bar{X} = 13.8$

so, the arithmetic mean is away from the center of data, compare with geometric mean.

(iii) Harmonic Mean

Is the reciprocal of arithmetic mean of the reciprocal of the given items.

$$\bar{H} = \frac{\sum f}{\sum \frac{f}{x_i}}$$

f is frequency of x_i

For $f=1$;

$$\bar{H} = \frac{n}{\sum \frac{1}{x_i}}$$

Ex what is the average flow rate, for a pump delivering 100 gal at 50 gpm, and 80 gal at 10 gpm.

gpm = $\frac{\text{gal}}{\text{min}}$

Sol- For the rate problem and speed problem, use the harmonic mean.

$$\bar{H} = \frac{2}{\frac{1}{50} + \frac{1}{10}} = 16.7 \text{ gpm}$$

Ex Aeroplane flies around square, which cover at speed of 100 km/hr one side, 200 km/hr the second side, 300 km/hr the third and 400 km/hr the fourth. what is the average speed?

$$\bar{H} = \frac{4}{\frac{1}{100} + \frac{1}{200} + \frac{1}{300} + \frac{1}{400}} = 192 \text{ km/hr}$$

Arithmetic mean $\bar{X} = \frac{100+200+300+400}{4} = \frac{1000}{4} = 250 \text{ km/hr}$

... gives unsuitable value. (355)

(iv) Median

For x_1, x_2, \dots, x_n , and these value arranged
as $x_1 \geq x_2 \geq \dots \geq x_n$ ascending order
or $x_1 \leq x_2 \leq \dots \leq x_n$ descending order

then the median is the value that divided x_1, x_2, \dots to two equal parts.

Ex 6, 10, 15, 18, 20 ترتيب

$$\text{median} = 15$$

Ex 6, 10, 15, 18, 20, 31

$$\text{median} = \frac{15 + 18}{2} = \frac{33}{2} = 16.5$$

Ex Find the median for the below samples

14, 15, 3, 4, 11, 7, 13, 12, 8, 5

L 1st arrange these data

3, 4, 5, 7, 8, 11, 12, 13, 14, 15

$$\text{median} = \frac{8 + 11}{2} = 9.5$$

When items are grouped frequency distribution, then the median is :-

$$M = l + \left(\frac{N}{2} - m \right) \frac{c}{f}$$

where :- N = total no. of item, and median lies in class interval of $(N/2)$ in cumulative frequency column.

l = lower limit of this class of length c ,

m = cumulative frequency upto this lower limit l .

f = frequency of this class

Ex Calculate the median for the following data relating to the number of automobiles crossing a point during an hour.

Time (min)	0-10	10-20	20-30	30-40	40-50	50-60
Number	14	17	22	26	23	18

Sol.

Time (min)	Frequency	Cumulative Frequency
0-10	14	14
10-20	17	31
20-30	22	53
30-40	26	79
40-50	23	102
50-60	18	120

where: $N = 120$, $\frac{N}{2} = \frac{120}{2} = 60$

So, 60 between 53-79 in cumulative freq. column and this corresponds to the class 30-40

- lower limit of this class $l = 30$
- length of class $c = 10$
- Cumulative freq. of this lower limit is $m = 53$
- Frequency of this class $f = 26$

$$\therefore M = l + \left(\frac{N}{2} - m\right) \frac{c}{f} = 30 + (60 - 53) \frac{10}{26}$$

$$\therefore M = 32.7 \text{ min.}$$

(357)

(V) Mode :

Is the size which occurs most frequently. For example, 3, 3, 3, 2, 2, 5

then, mode = 3

العدد هو الأهمية التي تتكرر بنسبة عالية
أو أهمها الأكثر شيوعاً.

In the case of grouped frequency distribution, then, mode (M_0) is :-

$$M_0 = l + \frac{cf_2}{f_1 + f_2}$$

l : lower limit of class of length c .

f_1, f_2 : frequency of preceding and succeeding classes respectively.

Also, there is an empirical formula to be used when the median (M), and mean \bar{X} are known.

$$M_0 = 3M - 2\bar{X}$$