Statistics

STATISTICS: It is the science of collection, presentation, analysis and interpretation of numerical data.

Data: It is the information in the form of numerical figures or a set of given facts.

(i) Ungrouped Data: Data obtained from direct observation is called raw data.

> eg: The markes obtained by 10 students in a Math test are: 5, 10, 6, 2, 1, 9, 8, 7, 3, 4

RANGE: The difference between the highest and Lowest number of data is called Range.

$$\text{Range} = 10 - 1$$
$$\text{Range} = 9 \qquad (\text{Refer above eg.})$$

OBSERVATION: Each numerical figure in a data is called observation.

FREQUENCY: The number of times a particular observation occurs is called the frequency.

(ii) Grouped Data :

To present the data in a more meaningful way, we condense the data into convenient number of classes or groups.

Class	Frequency
10-20	5
20-30	8
30-40	9
40-50	10
50-60	6
60-70	2

(a) CLASS INTERVAL: The group used to classify the data is called the class interval i.e. 10-20, 20-30.

(b) CLASS LIMITS : Each class is bounded by two figures, which are called the class limits.

- Upper Limit : In each class interval, the R.H.S. or the greatest number is the upper-class limit. i.e. 10-20 ↑ upper limit

- Lower limit : In each class interval, the L.H.S. or the smallest number is the lower-class limit. i.e. 10-20 ↑ Lower limit

(C) CLASS SIZE : It is the difference between the upper limit and lower limit.

eg: 10-20 is class interval
lower limit = 10, upper limit = 20
Class size = 20-10
= 10

(d) CLASS MARK : The mid point of each class interval is the class mark.

$$\text{Class mark} = \frac{\text{Upper limit} + \text{Lower limit}}{2}$$

i.e. the class mark of 20-30 is $\frac{1}{2}(30+20)$
$= 25$

(e) Number of Class Interval $= \dfrac{\text{Range}}{\text{Class size}}$

TWO METHODS OF CLASSIFYING DATA according to the Class Intervals:
(i) Exclusive Method
(ii) Inclusive Method

(i) EXCLUSIVE METHOD: When the class interval are so fixed that the upper limit of one class is the lower limit of the next class, it is known as exclusive method.

eg:

Wages (Rs.)	No. of Workers
1000 - 1100	125
1100 - 1200	150
1200 - 1300	200
1300 - 1400	250

(ii) INCLUSIVE METHOD : In this method the classes are so formed that the upper limit of a class is included in that class.

eg :

Wages (Rs)	No. of Workers
1000 – 1099	125
1100 – 1199	150
1200 – 1299	200
1300 – 1399	250

GRAPHICAL REPRESENTATION OF STATISTICAL DATA:

- Bar Graphs
- Histogram
- Frequency Polygon

1) BAR GRAPH:

- The width of the bars should be uniform throughout.
- The gap between one bar and another should be uniform throughout.
- Bars may be either horizontal or vertical.

2.) HISTOGRAM:

- Grouped data is represented graphically by histograms.
- No gap between any two successive rectangles.
- The class interval are to be taken along an x-axis
- The height represents the frequencies of the respective class intervals.

Construction Of HISTOGRAM:

2.1 Construction of a histogram of a continuous grouped frequency distribution with equal class interval:

eg:

Marks:	0-10	10-20	20-30
No. of Students (Frequency)	4	10	8

Kink Sign (⟋\/⟍): It is a zig-zag line that is usually drawn in the x-axis near the origin when the scale on the axis does not start from zero.

2.2. **Construction of a histogram of a continuous grouped frequency distribution with unequal class intervals:**

- Choose a suitable scale along X-axis and represent class-limits on it.

- Determine the class-interval which has the minimum class size. Let the minimum class size be 'h'.

- Compute the adjusted frequencies of each class by using the following formula:

$$\text{Adjusted frequency of a class} = \frac{h}{\text{class size}} \times \text{frequency of the class}$$
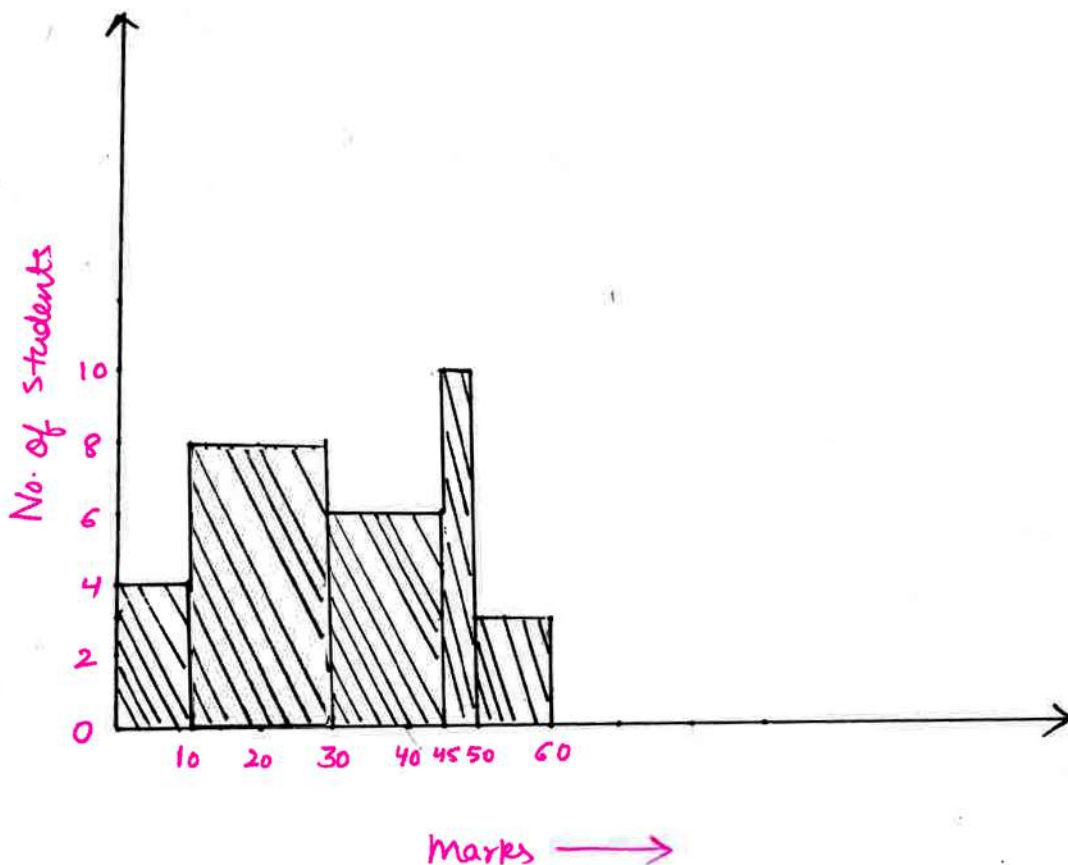
eg:

Marks :	0-10	10-30	30-45	45-50	50-60
No. of Students : (Frequency)	8	32	18	10	6

here, minimum class size = 5

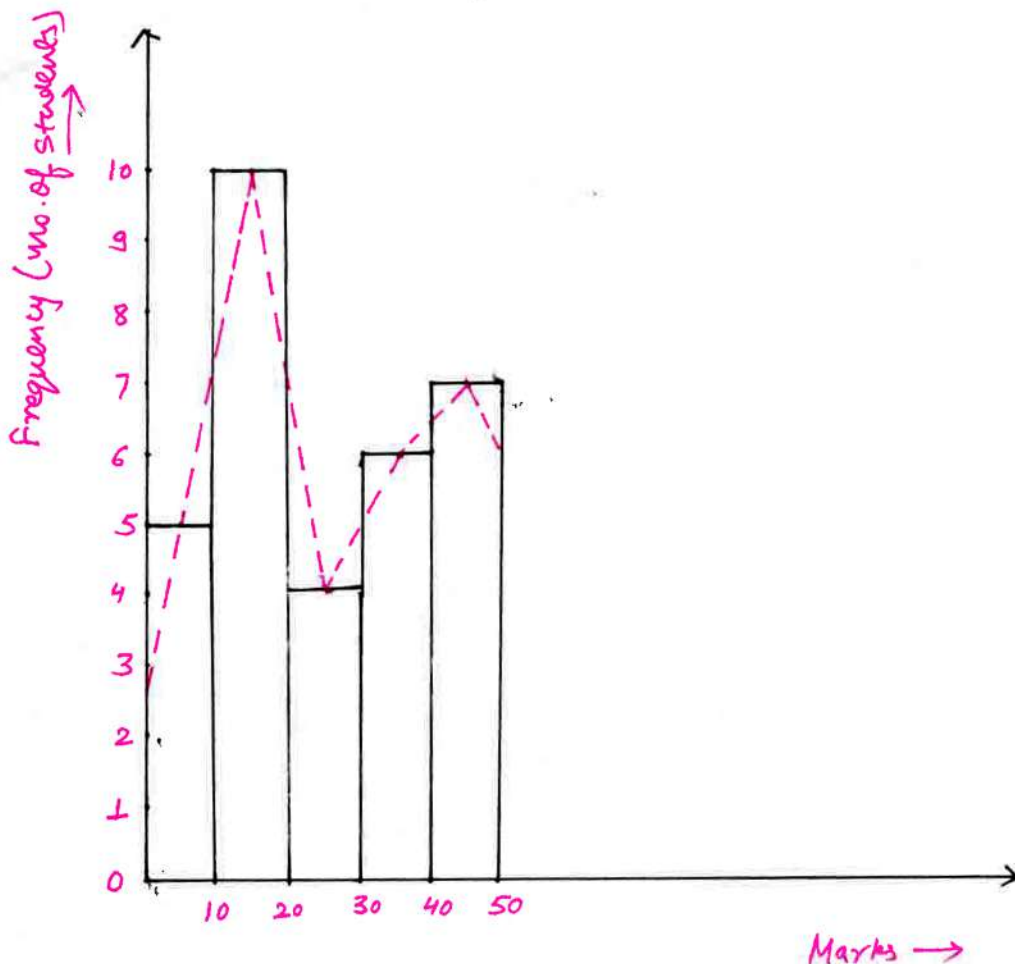| Class-Intervals (Marks)	No. of Students (Frequency)	Adjusted Frequency
0-10	8	$\frac{5}{10} \times 8 = 4$
10-30	32	$\frac{5}{20} \times 32 = 8$
30-45	18	$\frac{5}{15} \times 18 = 6$
45-50	10	$\frac{5}{5} \times 10 = 10$
50-60	6	$\frac{5}{10} \times 6 = 3$

FREQUENCY POLYGON: Frequency Polygon is another method of representing frequency distribution (grouped-data) graphically.

(i) Frequency Polygon by using Histogram:

eg:

Marks:	0-10	10-20	20-30	30-40	40-50
Frequency:	5	10	4	6	7

(ii) Frequency Polygon without using Histogram:

- Compute the mid points of class-intervals i.e. class marks.

- Represent the class-mark on X-axis on a suitable class scale.

- Represent the frequencies on Y-axis on a suitable scale.

- Plot the points (x_i, f_i) where x_i denotes class marks and f_i corresponding frequency.

- Join the points.

ex:

Age (in years)	0-2	2-4	4-6	6-8	8-10
Frequency	2	4	6	8	2