BCA, SECOND SEMESTER

CCSIT, TMU

DATA AND ITS REPRESENTATION

DR. PAVAN KUMAR SINGH
ASSISTANT PROFESSOR
DEPARTMENT OF PHYSICS
FACULTY OF ENGINEERING,
TEERTHANKER MAHAVEER UNIVERSITY

SYLLABUS

UNITS	Contents	
Unit-1:	**Kinds of Data, Collection of Data, Representation of Data: One Dimensional, Two-Dimensional, Pie Diagram, Graphical Representation of Data: Histograms, Frequency Polygons, Frequency Curve, Cumulative Frequency Curve.**	**8 Hours**
Unit-2:	**Measures of Central Tendency: Mean (Direct method, short cut method, Step Deviation method), Median, Mode, Relation between Mean, Median and Mode.**	**8 Hours**
Unit-3:	**Measures of Dispersion: Range, Coefficient of Range, Mean Deviation, Coefficient of Mean Deviation, Quartile Deviation, Coefficient of Quartile Deviation, Standard Deviation.**	**8 Hours**
Unit-4:	**Probability: Random Experiment, Sample Space, Event, Types of Events, Definition of probability, Addition theorem of Probability, Multiplication theorem of probability.**	**8 Hours**
Unit-5:	**Probability Distribution: Binomial distribution, Mean and Variance of Binomial distribution, Poisson Distribution, Mean and Variance of Poisson Distribution.**	**8 Hours**

Statistics

- **Statistics is the branch of mathematics which deals with collecting, analyzing and interpreting data.**
- **Statistics can be used to predict the future, determine the probability that a specific event will happen, or help answer questions about a survey.**
- **Statistics is used in many different fields such as business, medicine, biology, psychology and social sciences.**

Functions of Statistics

1. Collection of data
2. Tabulation of data
3. Analysis of data
4. Interpretation of data

Data

- **Data is a collection of facts and figures gathered for a specific purpose, such as conducting surveys or performing analyses.**

- **Simply, Data can be defined as a systematic record of a particular quantity**

- When organized in an orderly manner, it becomes **information**.

Types of Data

1. **Qualitative Data**:
    1. Qualitative data, represents characteristics or attributes.
    2. These attributes can be observed but cannot be precisely computed or calculated.
    3. Examples include data related to **intelligence, honesty, wisdom, cleanliness, and creativity**.
    4. Qualitative data is more exploratory in nature and is often analyzed using **grouping methods** and visual representations like **pie charts**.

2. Quantitative Data:

1. Quantitative data represents numerical value.
2. Examples include data on the **number of students playing different sports** in a class.
3. Quantitative data is often analyzed using tools like **line graphs**, **scatter plots**, or **probability density functions**.

Data Collection

- Depending on the source, it can classify as primary data or secondary data.
- **1. Primary Data**
- Primary data refers to information collected directly by an investigator for a specific purpose.
- It is considered “pure” because no statistical operations have been performed on it yet.
- An example of primary data is the **Census of India**, where data is collected afresh.

- **2. Secondary Data**:
- Secondary data comes from existing sources where it has already been collected by other researchers or investigators.
- This data may be available in published or unpublished forms.
- It is considered “impure” because statistical operations may have been applied to it.
- Examples include data from **government websites, repositories, books, and journals**

Graphical representation of data

- A graphical representation is a visual display of data and statistical results.
- It is more effective than presenting the data in tabular form. Graphical representation is another way of analysing numerical data.
- A graph is a chart through which data are represented in the form of lines or curves drawn across the coordinated points plotted on the surface.

Advantages

- Easlily understood, easy intrepretation. More attractive.
- Useful in comparing relationship between two or more sets of figures.
- Enable statistical problems in a visual form.

Types

- Line diagram/graph
- Bar diagram
- Pie chart
- Histogram
- Frequency polygon
- Ogive/commulative frequency polygon.

Graphical representation of ungrouped data

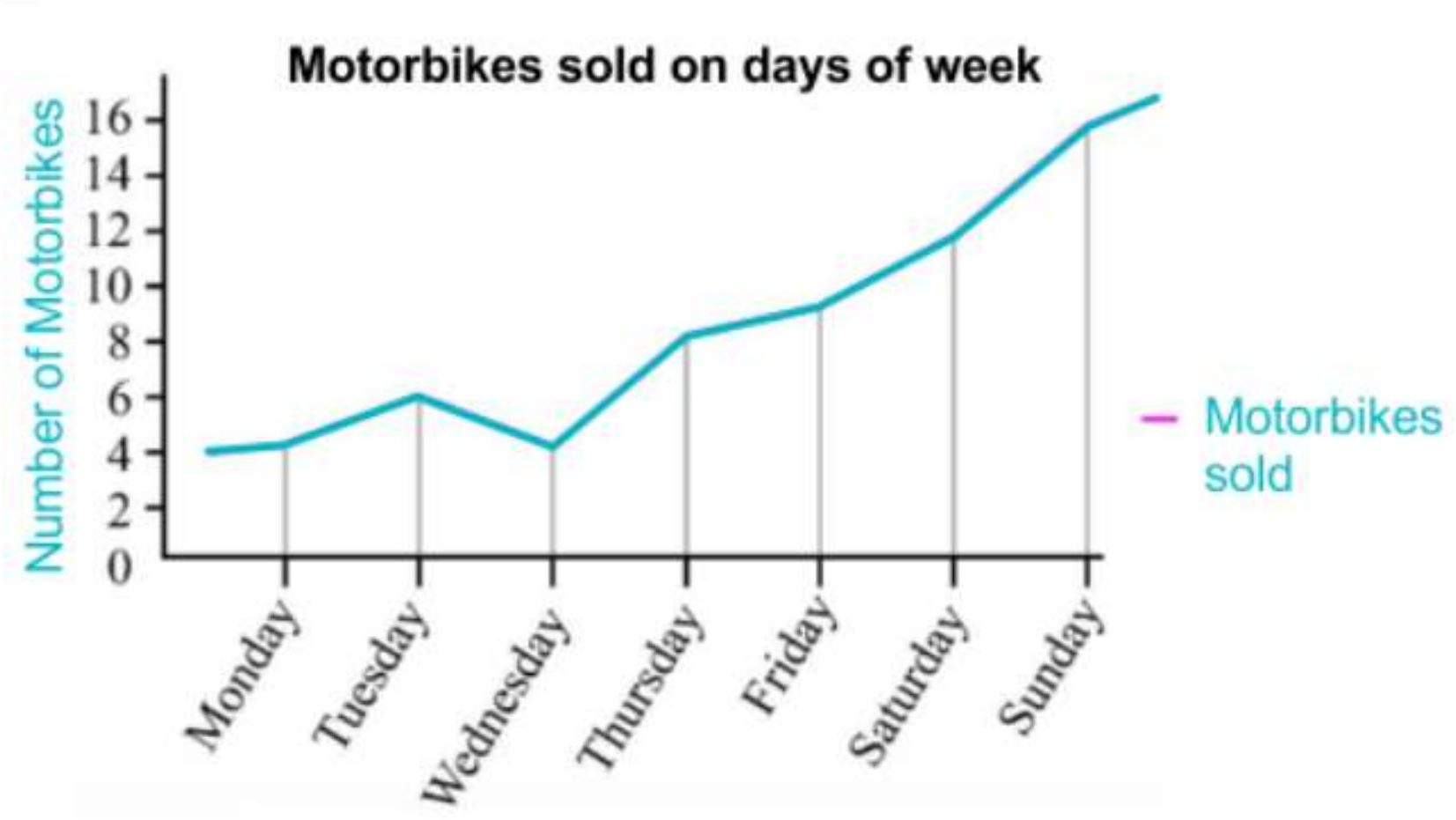
- can be represented by using
  1.Line graph.
  2. Bar graph/bar diagram
  3. Circle graph/pie diagrams
  4. pictograms

Line graph/graph

- A line graph is a type of graph which is used to show information that changes over time. We can plot the line graph by joining several points with straight lines.
- A line graph is also called a line chart.
- A line graph contains two axes i.e., the x-axis and the y-axis.

Line graph

Example: The given figure is an example of a line graph.



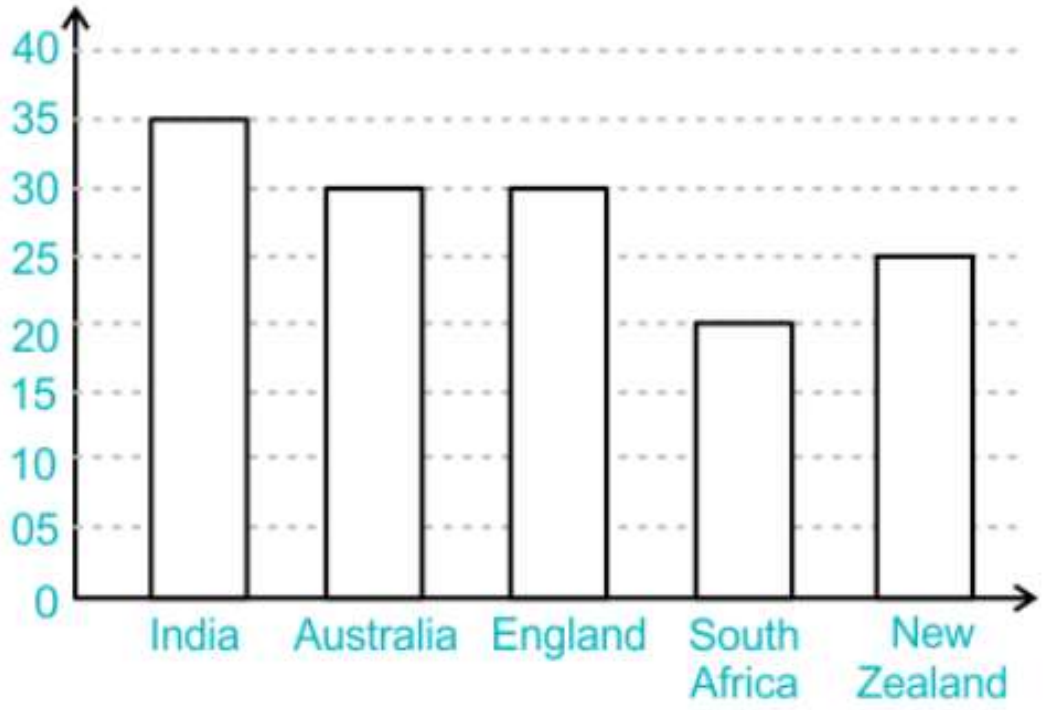
- Most common graphical representation
- By plotting X axis on horizontally while Y axis vertically.
- Find out the intersecting point or origin and join all intersections.
- Example: Motorbikes sold on days of week.

BAR GRAPH/DIAGRAM

- **Bar Graph:** A BAR GRAPH is a pictorial representation of data in which bars of uniform width are drawn with equal spacing between them. The height of each bar depends on the values they represent.

EXAMPLE OF BAR GRAPH

- Example: The given bar graph shows the number of matches played by different teams.



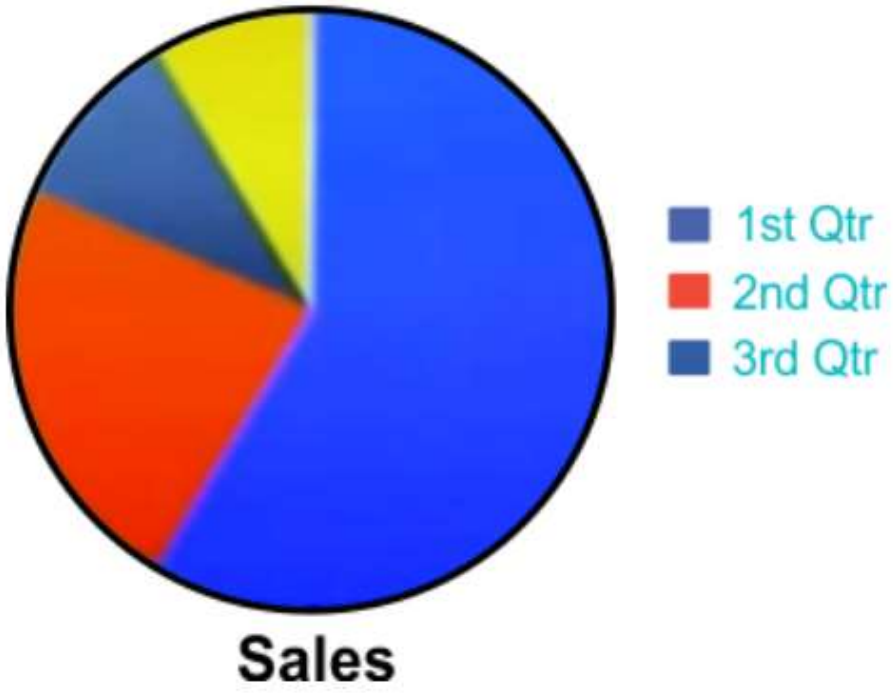
Pie Chart/Diagram

- A pie chart is a representation of values as slices of a circle with different colours. The slices are labelled and the numbers corresponding to each slice are also represented in the chart.

Pie Chart

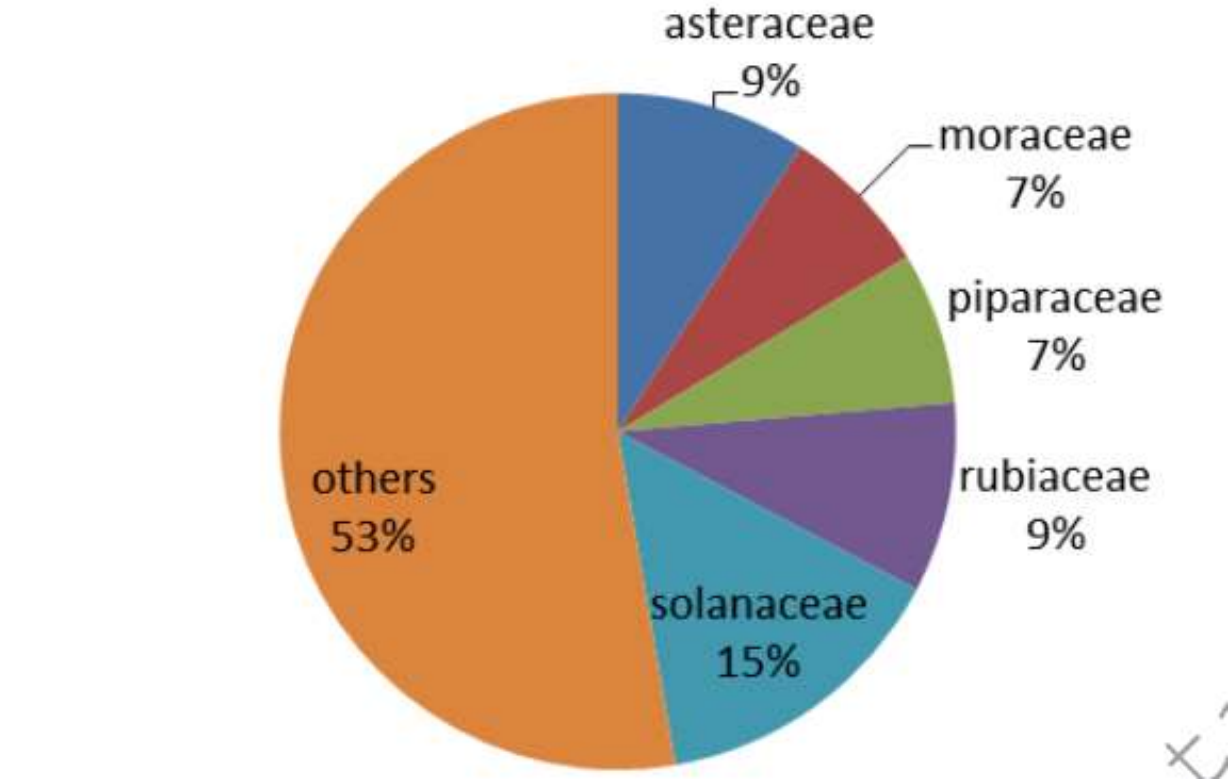
- Is a circular graph that represent total value in circle and components in part wise.
- Useful in comparing components and total value.
- Data is expressed in percentage of total value.

Example: Annual sales of a company.



Asteraceae	5
Moraceae	4
Piperaceae	4
Rubiaceae	5
Solanaceae	8
Others	29

$$Angle = \frac{Value\ of\ one\ component}{Total\ of\ all\ components} \times 360$$



Pie Chart Example:

Marks obtained in test examination by a B.SC. (PCM) students of Moradabad.

PHYSICS	MATHEMATICS	CHEMISTRY
65	60	55

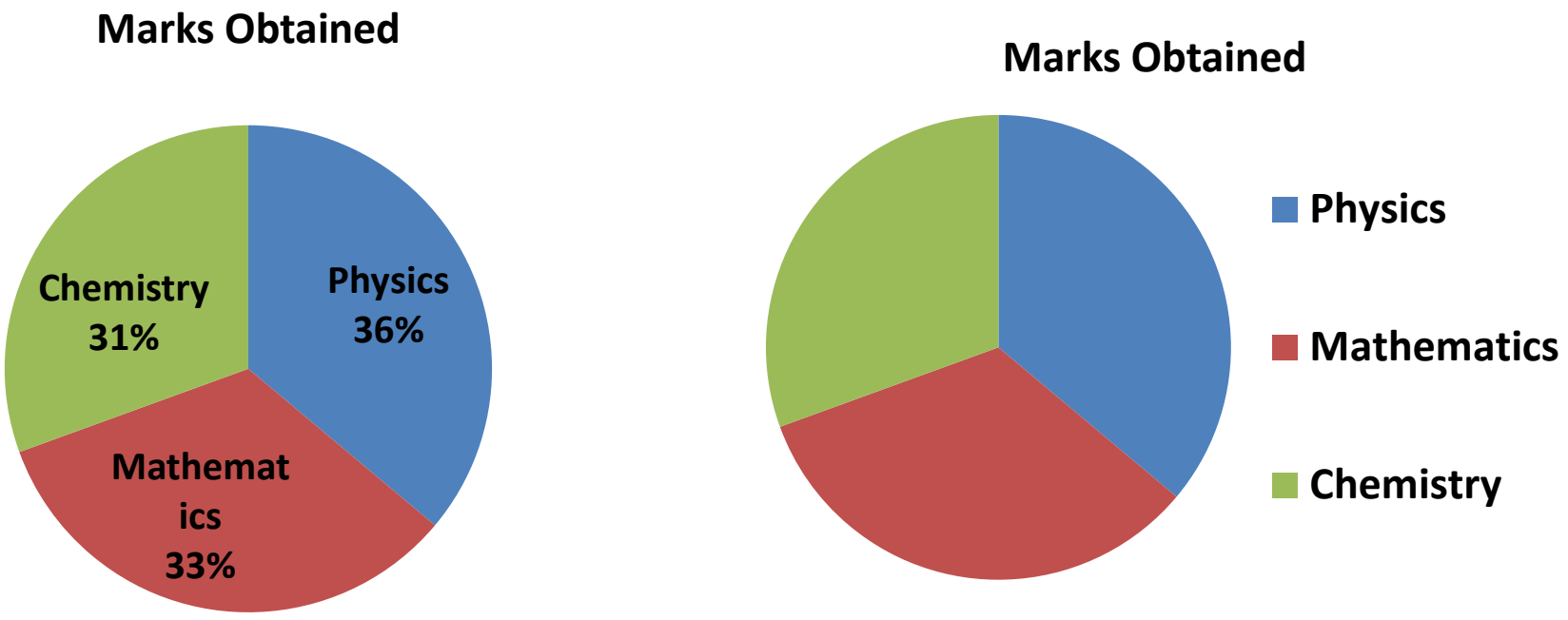


Fig. Pie Chart showing marks of three subjects

Pie Chart: Angle calculation

SUBJECT	MARKS OBTAINED	ANGLE
PHYSICS	65	130
MATHEMATICS	60	120
CHEMISTRY	55	110

Total Number of Components=60+65+55=180

$$Angle = \frac{Value\ of\ one\ component}{Total\ of\ all\ components} \times 360$$

Graphical representation of grouped data

Can be represented by using:

- Histogram
- Frequency polygonal
- Cumulative frequency curve or ogive
- Ratio diagram/ arithlog

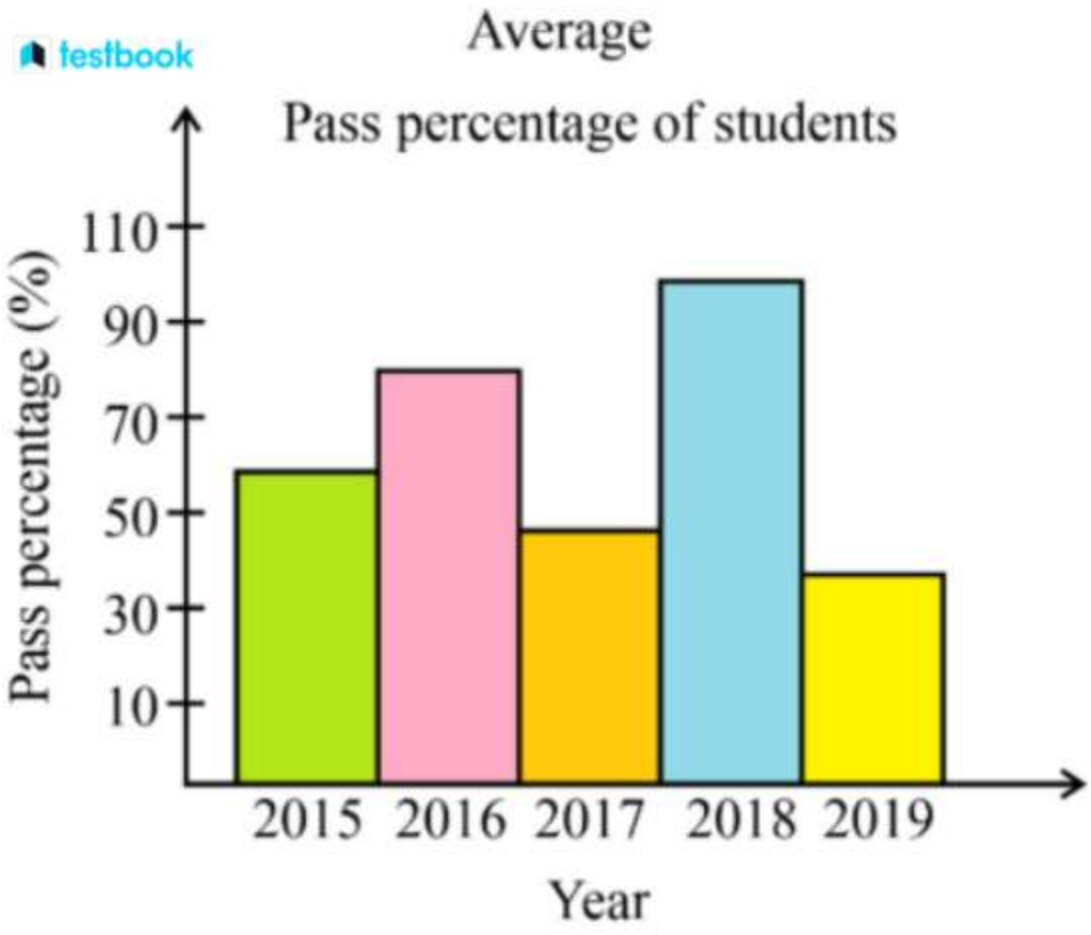
Histogram

- The histogram is a type of graph that shows continuous data represented on a two-dimensional graph. In the histogram, the bars are placed side by side without a gap between consecutive bars. The histogram is a very popular method of presenting continuous series and class intervals.

Histogram

- Histogram is represented by a rectangular bar to depict frequency distribution.
- Size of the class interval is represented by width
- Size of the frequency is represented by height.
- Class boundaries/intervals is important in the construction of histogram and represent in horizontal or X axis of the graph.
- Frequency is represented as height in the graph on Y axis.
- Histogram is essentially an area diagram composed of series of adjacent ractangles.

Example: The given figure is an example of a histogram.



Frequency Polygons

- A curve obtained by joining the middle of the histograms starts from (highest) histogram to immediate lower histogram successively and form into a figure with a many angles or polygon.
- It is used, where class intervals are equal with discrete (not joined) variables.
- It gives idea about the shape of the frequency distribution.

Example 1: Construct a histogram and frequency polygon for the following data:

100-150	150-200	200-250	250-300	300-350
4	6	13	5	2

SOLUTION: THIS IS THE CASE OF EQUAL CLASS INTERVAL.

CLASS INTERVAL	**FREQUENCY**	**COMMULATIVE FREQUENCY (C.F.)**
100-150	4	4
150-200	6	10
200-250	13	23
250-300	5	28
300-350	2	30

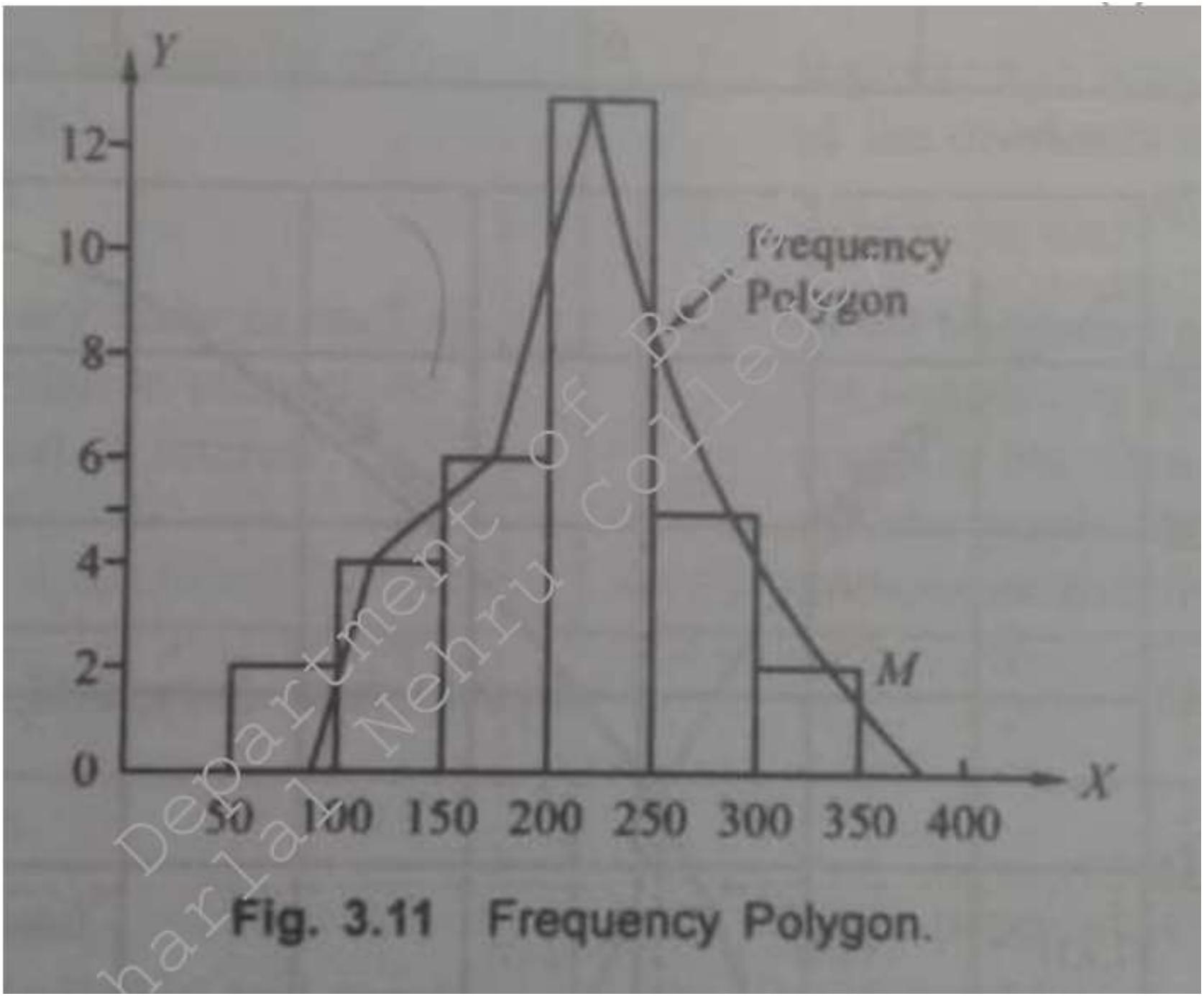


Fig. 3.11 Frequency Polygon.

***Example* 2.6.** The table below gives the distribution of the age of women at the time of marriage in Sri Lanka.

Age groups (years)	*No. of women*
15 – 19	11
20 – 24	36
25 – 29	28
30 – 34	13
35 – 39	7
40 – 44	3
45 – 49	2

The given distribution is not continuous as the upper limit of the preceding class is not the lower limit of the following class. Hence, it is smoothened. The difference between 20 and 19 is 1. Therefore, 0.5 is to be subtracted from the lower limit of the classes and 0.5 is to be added to the upper limit of all classes. Since, the difference is constant, the same quantity is subtracted and added in all classes.

Smoothened age groups (years)	*No. of women*
14.5 – 19.5	11
19.5 – 24.5	36
24.5 – 29.5	28
29.5 – 34.5	13
34.5 – 39.5	7
39.5 – 44.5	3
44.5 – 49.5	2

Cumulative Frequency Distribution

It can be formed on "less than" or "more than" basis. In example 2.5, we can either make the distribution on the number of children with a birth weight less than a particular weight or the number of children with a birth weight more than a specified weight. The cumulative frequency distributions for the data given in example 2.5 are presented below:

Table 2.2: Cumulative frequency and percentage distributions

Less than type		*Percentage*	*More than type*		*Percentage*
Birth weight	*cu.fr.*		*Birth weight*	*cu.fr.*	
Less than 2.4	5	17	2.0 or more	30	100
Less than 2.8	10	33	2.4 or more	25	83
Less than 3.2	19	63	2.8 or more	20	67
Less than 3.6	23	77	3.2 or more	11	37
Less than 4.0	27	90	3.6 or more	7	23
Less than 4.4	30	100	4.0 or more	3	10

Note: It must be remembered that the cumulative frequency of a less than type frequency distribution always refers to the upper limit of the class interval and for more than type it refers to the lower limit of the class interval.

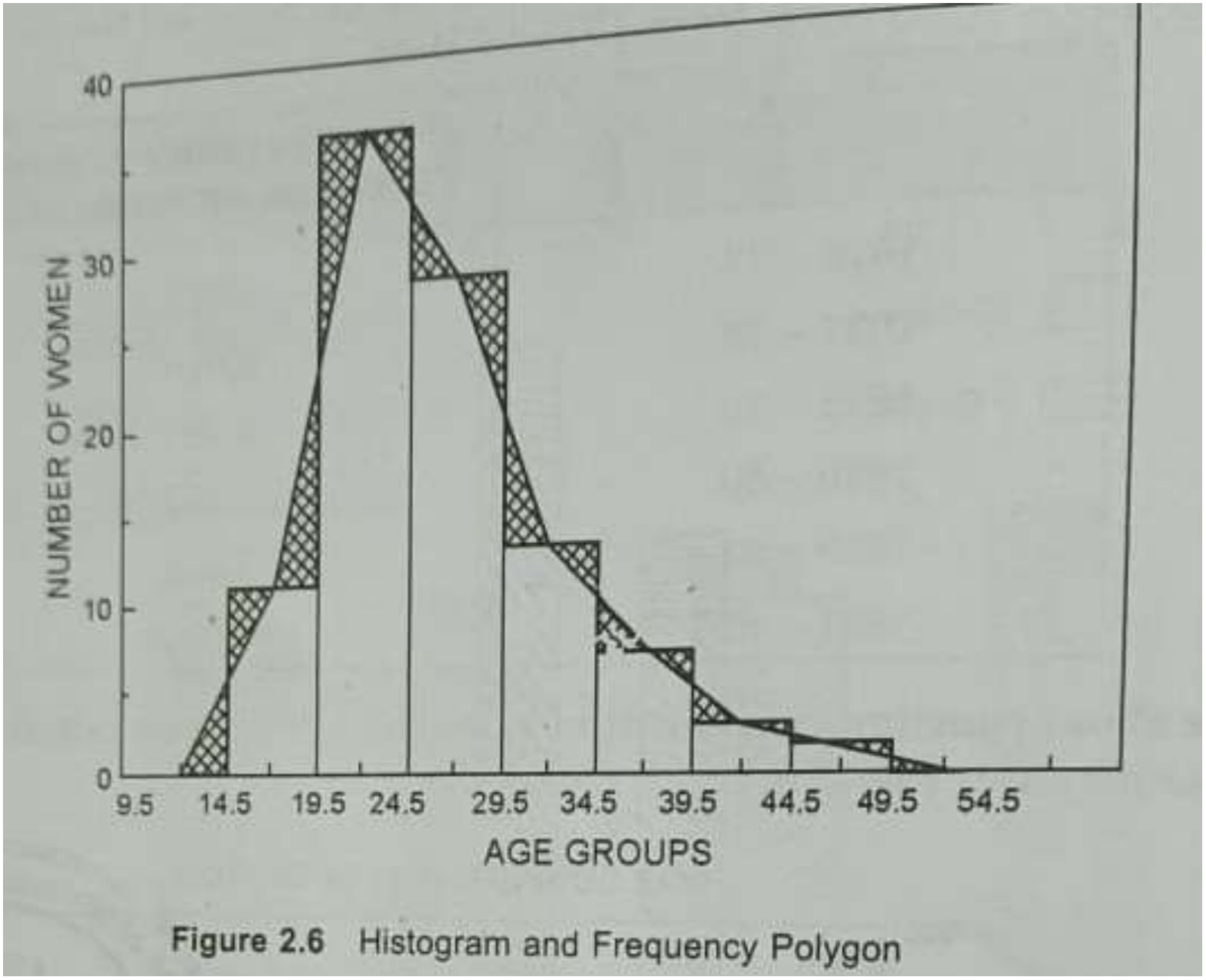
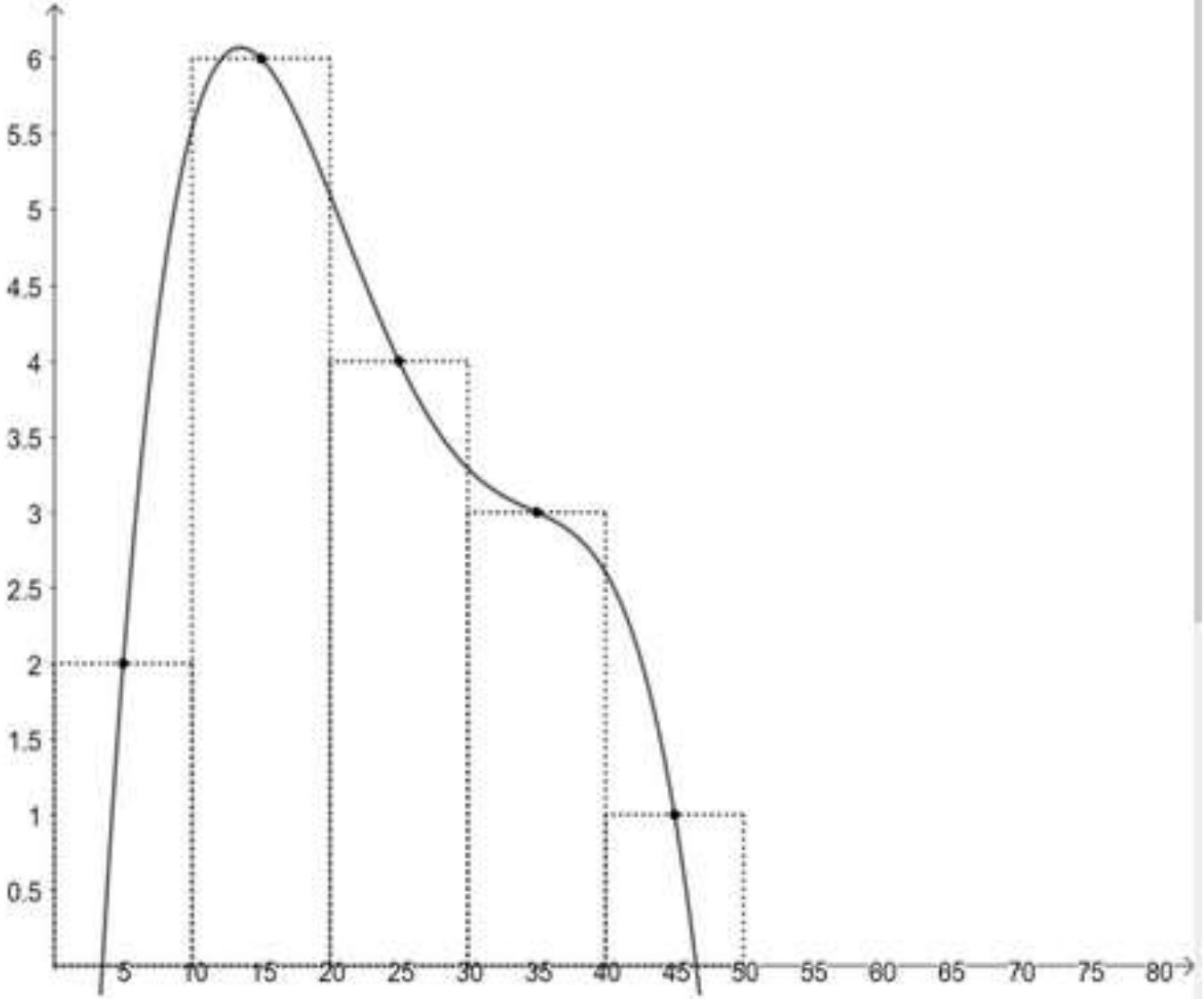


Figure 2.6 Histogram and Frequency Polygon

Frequency Curve

- A frequency curve is **a limiting form of a histogram or a frequency polygon**. A frequency curve for a given distribution can be obtained by drawing a smooth, free hand curve through the midpoints of the upper sides of the rectangles forming the histogram.


6
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80

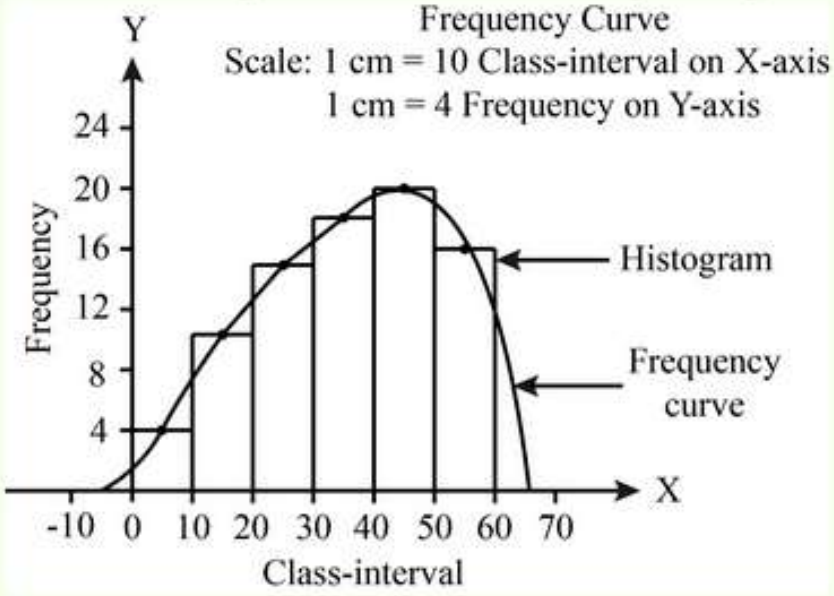
Question

Draw a frequency curve for the following distribution.

Class-interval	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
Frequency	4	10	15	18	20	16

Solution

The frequency distribution is continuous with equal class-intervals. We will first prepare a histogram and then the frequency curve. The class-intervals are shown along the X-axis and frequency on the Y-axis.



Question

Draw a cumulative frequency curve and cumulative frequency polygon for the following frequency distribution by less than method.

Age (in years)	0-9	10-19	20-29	30-39	40-49	50-59	60-69
No. of Persons:	5	15	20	23	17	11	9

Solution

The given frequency distribution is not continuous. So, we first make it continuous and prepare the cumulative frequency distribution as under.

Age (in years)	Frequency	Age less than	Cumulative frequency
-0.5-9.5	5	9.5	5
9.5-19.5	15	19.5	20
19.5-29.5	20	29.5	40
29.5-39.5	23	39.5	63
39.5-49.5	17	49.5	80
49.5-59.5	11	59.5	91
59.5-69.5	9	69.5	100

Now, we plot points

(9.5, 5), (19.5), (29.5, 40), (39.5, 63), (49.5, 80), (59.5, 91)and(69.5, 100). and join them by a free hand smooth curve to obtain the required ogive as shown in Fig. 7.5.

The cumulative frequency polygon is obtained by joining these points by line segments as shown in Fig. 7.6.

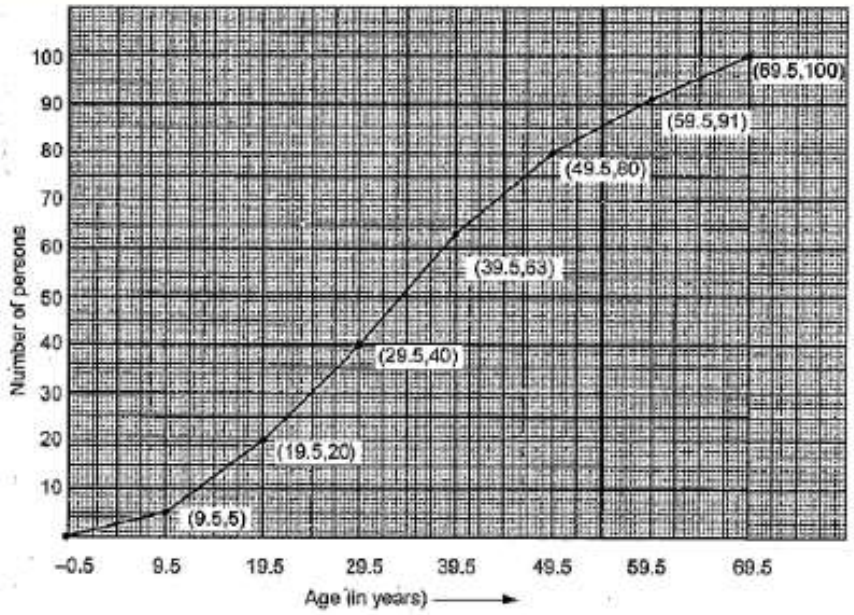


Fig. 7.5 Cumulative frequency curve

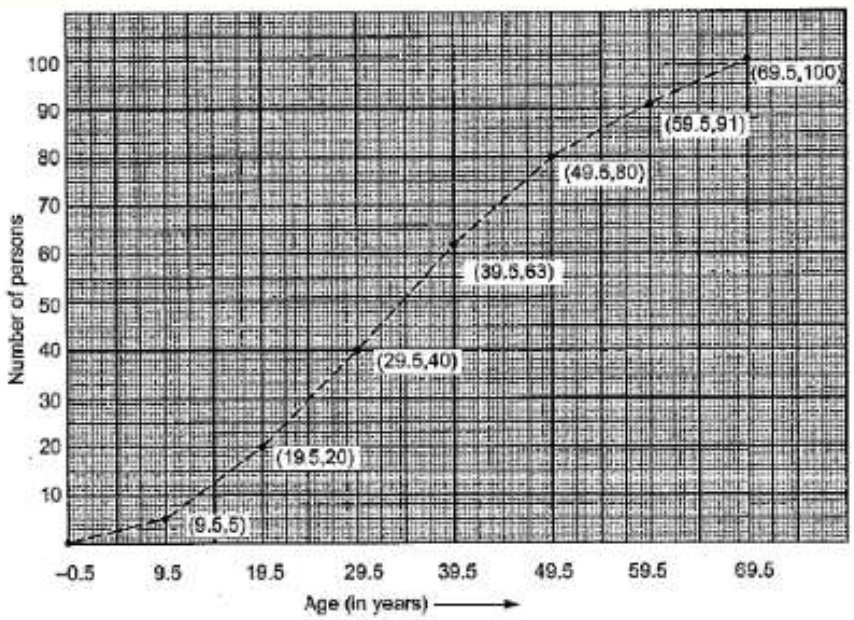


Fig. 7.6 Cumulative frequency polygon

Ogives (cumulative frequency curves)

- The Ogive is defined as the frequency distribution graph of a series. The Ogive is a graph of a cumulative distribution, which explains data values on the horizontal plane axis and either the cumulative relative frequencies, the cumulative frequencies or cumulative per cent frequencies on the vertical axis.
- Cumulative frequency is defined as the sum of all the previous frequencies up to the current point. To find the popularity of the given data or the likelihood of the data that fall within the certain frequency range, Ogive curve helps in finding those details accurately.
- Create the Ogive by plotting the point corresponding to the cumulative frequency of each class interval. Most of the Statisticians use Ogive curve, to illustrate the data in the pictorial representation. It helps in estimating the number of observations which are less than or equal to the particular value.

Ogive Graph

The two methods of Ogives are:

- Less than Ogive
- Greater than or more than Ogive

Less than Ogive:
The frequencies of all preceding classes are added to the frequency of a class. This series is called the less than cumulative series. It is constructed by adding the first-class frequency to the second-class frequency and then to the third class frequency and so on. The downward cumulation results in the less than cumulative series.

Greater than or More than Ogive

- The frequencies of the succeeding classes are added to the frequency of a class. This series is called the more than or greater than cumulative series. It is constructed by subtracting the first class, second class frequency from the total, third class frequency from that and so on. The upward cumulation result is greater than or more than the cumulative series.

Ogive Example

Question 1:

Construct the more than cumulative frequency table and draw the Ogive for the below-given data.

Marks	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80
Frequency	3	8	12	14	10	6	5	2

Solution:

"More than" Cumulative Frequency Table:

Marks	Frequency	More than Cumulative Frequency
More than 1	3	60
More than 11	8	57
More than 21	12	49
More than 31	14	37
More than 41	10	23
More than 51	6	13
More than 61	5	7
More than 71	2	2

Plotting an Ogive:

Plot the points with coordinates such as (70.5, 2), (60.5, 7), (50.5, 13), (40.5, 23), (30.5, 37), (20.5, 49), (10.5, 57), (0.5, 60).

An Ogive is connected to a point on the x-axis, that represents the actual upper limit of the last class, i.e.,(80.5, 0)

Take x-axis, 1 cm = 10 marks

Y-axis = 1 cm – 10 c.f

More than the Ogive Curve: