Statistics

- **Construction of histogram of different class intervals**
  A histogram is a graphical representation of grouped data.

  Example: Represent the given data in the form of a histogram.

 | Marks | 0 – 20 | 20 – 30 | 30 – 50 | 50 – 60 |
  |---|---|---|---|---|
  | Number of students | 10 | 13 | 16 | 11 |

  Solution: Here, the minimum class size is 10.
  Thus, we can find the adjusted frequencies using the following formula,

  $$\text{Adjusted frequency of a class} = \frac{\text{Minimum class size}}{\text{Class size}} \times \text{Frequency}$$

 | Marks | Number of students | Class size | Adjusted frequency |
  |---|---|---|---|
  | 0 – 20 | 10 | 20 | $\frac{10}{20} \times 10 = 5$ |
  | 20 – 30 | 13 | 10 | $\frac{10}{10} \times 13 = 13$ |
  | 30 – 50 | 16 | 20 | $\frac{10}{20} \times 16 = 8$ |
  | 50 – 60 | 11 | 10 | $\frac{10}{10} \times 11 = 11$ |

  The histogram of the given data can be drawn as follows:
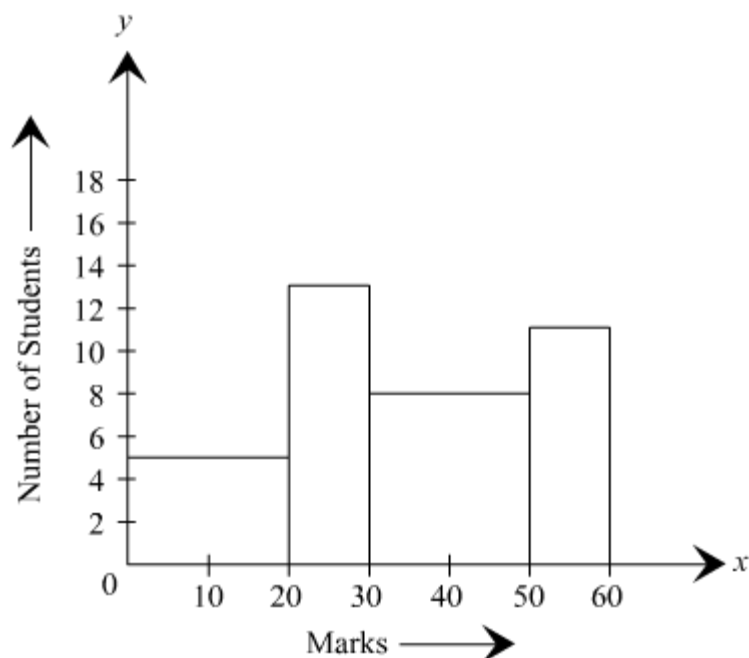
- **Interpretation of histograms**

  The interpretation of data represented by a histogram can be done by looking at it.

  Example: The following histogram shows the production of food grains (in thousand tonnes) over a period of time.
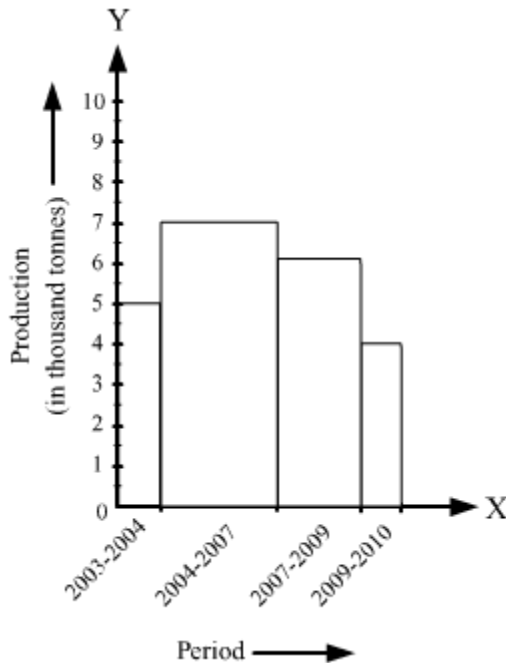
  (i) What is the total production of food grains from 2004 to 2009?

  (ii) In which periods were the production of food grains the highest and the lowest?

  Solution:

  (i) The total production of food grains from 2004 to 2009 can be ascertained by adding the heights of the class intervals 2004–2007 and 2007–2009.
  ∴ Total production of food grains from 2004 to 2009 = 7000 tonnes + 6000 tonnes = 13000 tonnes

  (ii) It is clear from the histogram that the bar corresponding to the class interval 2004–2007 is the tallest, and that corresponding to the class interval 2009–2010 is the shortest. So, the production of food grains was the highest in the period 2004–2007 and the lowest in the period 2009–2010 (i.e., 7000 tonnes and 4000 tonnes respectively).

- **Construction of frequency polygons**

  A frequency polygon is a continuous curve obtained by plotting and joining the ordered pairs of class marks and their corresponding frequencies.

There are two ways to construct a frequency polygon.

o   The frequency polygon for a grouped data is drawn by first drawing its histogram and then by joining the mid-points of the top of bars and the mid-points of the classes preceding and succeeding the lowest and highest class respectively.

o   One other way of drawing a frequency polygon is by plotting and joining the ordered pairs (of class marks and their corresponding frequencies) with the mid-points of the classes preceding and succeeding lowest and highest class respectively.

**Example**: Here are the weights (in kg) of the babies born in a hospital during a particular week.

2.3, 2.0, 2.5, 2.7, 3.0, 3.2, 3.1, 2.2, 3.0, 2.5, 2.4, 3.0, 2.3, 2.4, 2.8
Draw a histogram for the data and then draw a frequency polygon using it.

Solution: The frequency distribution table of the given data is as follows:

Class interval	Frequency
2.0–2.5	6
2.5–3.0	4
3.0–3.5	5

The histogram and frequency polygon for the given data can be drawn as:
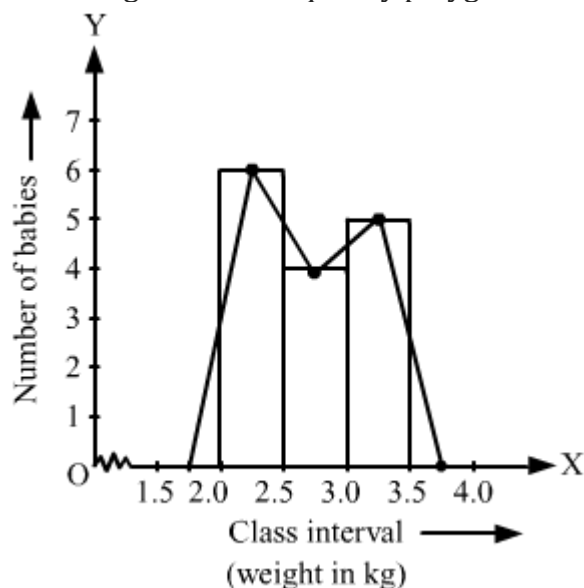
- **Mean of data sets**

  Mean or average of a data is given by the formula,

$$\text{Mean} = \frac{\text{Sum of all observations}}{\text{Number of observations}}$$

Note:

- Mean always lies between the highest and lowest observations of the data.
- It is not necessary that mean is any one of the observations of the data.

1. If the mean
of n observations $x_1, x_2, x_3 \ldots x_n$ is $\bar{x}$ then $(x_1 - \bar{x}) + (x_2 - \bar{x}) + (x_3 - \bar{x}) + \ldots + (x_n - \bar{x})$
$= 0$.

2. If the mean of n observations $x_1, x_2, x_3 \ldots x_n$ is $\bar{x}$ then the mean
of $(x_1 + p), (x_2 + p), (x_3 + p), \ldots, (x_n + p)$ is $(\bar{x} + p)$.

3. If the mean of n observations $x_1, x_2, x_3 \ldots x_n$ is $\bar{x}$ then the mean
of $(x_1 - p), (x_2 - p), (x_3 - p), \ldots, (x_n - p)$ is $(\bar{x} - p)$.

4. If the mean of n observations $x_1, x_2, x_3 \ldots x_n$ is $\bar{x}$ then the mean
of $px_1, px_2, px_3, \ldots, px_n$ is $p\bar{x}$.

5. If the mean of n observations $x_1, x_2, x_3 \ldots x_n$ is $\bar{x}$ then the mean
of $\frac{x_1}{p}, \frac{x_2}{p}, \frac{x_3}{p}, \ldots, \frac{x_n}{p}$ is $\frac{\bar{x}}{p}$.

Example: The runs scored by a batsman in 6 matches are as follows:
24, 126, 78, 43, 69, 86
What is the average run scored by the batsman?

Solution: Total number of runs scored = 24 + 126 + 78 + 43 + 69 + 86
= 426
Number of matches = 6
$\therefore$ Average runs scored = $\frac{426}{6} = 71$

- **Mean of a frequency distribution table**
  Formula to calculate mean from a frequency distribution table is given by,
  $$\text{Mean}(\bar{x}) = \frac{\sum f_i x_i}{\sum f_i}$$, where f_i and x_i are the frequencies and observations respectively.

  For example, consider the following data.

Marks (x_i)	Number of students (f_i)	$f_i x_i$
10	4	40
20	3	60
30	5	150

	40	7	280
	50	10	500
		$\sum f_i = 29$	$\sum f_i x_i = 1030$

Here, $$\text{Mean}(\bar{x}) = \frac{1030}{29} = 35.52 \; (\text{approx})$$

- **Median**
  Median is the value of the middlemost observation when the data is arranged in increasing or decreasing order.

  To find the median, the observations are arranged in ascending or descending order and then, if the number of observations (n) is odd, the value of $\left(\frac{n+1}{2}\right)^{th}$ observation is the median.
  If the number of observations (n) is even, then the mean of the values of $\left(\frac{n}{2}\right)^{th}$ and $\left(\frac{n}{2}+1\right)^{th}$ observations is the median.

  Example: The weights of 7 students are as follows: 30, 35, 41, 29, 28, 32, 30. What is the median of this data?

  Solution: The observations in ascending order are 28, 29, 30, 30, 32, 35, 41.
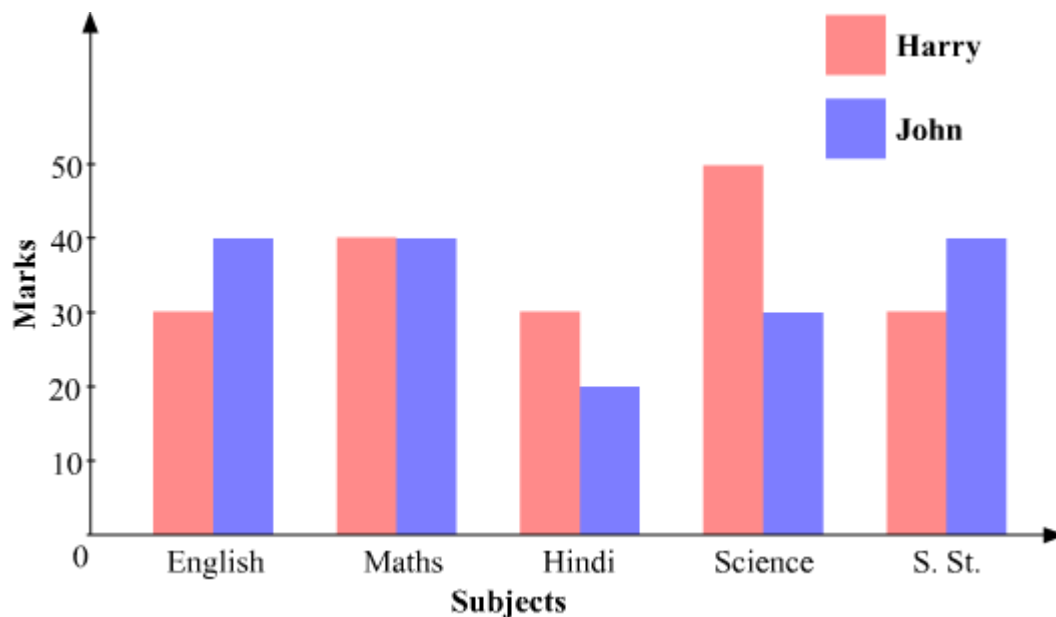 Here, n = 7 (which is odd)
  $$\therefore \text{Median} = \left(\frac{7+1}{2}\right)^{th} \text{observation}$$
  $$= 4^{th} \text{observation}$$
  $$= 30$$

- We can interpret the double bar graph by reading and analyzing it.

  Example: The given double bar graph represents the marks obtained by Harry and John in different subjects.

We can analyze the given double bar graph and answer the following questions.

1. What is the differnce between the marks scored by Harry and John in Science?

   Answer: Marks scored by Harry in science = 50
   Marks scored by John in science = 30

   Difference = 50 – 30 = 20

2. In which subject both the students got equal marks?

   Answer: Both the students got equal marks in Maths.

3. Who is better in studies among Harry and John?

   Answer: Total marks obtained by Harry = 30 + 40 + 30 + 50 + 30 = 180

   Total marks obtained by John = 40 + 40 + 20 + 30 + 40 = 170

   Total marks obtained by Harry is more than John. So, Harry is better in studies than John.

- In a pictograph, pictures of objects are used for representing data. Tally marks cannot be used for representing huge numbers. However, these numbers can be represented with the help of pictographs.

- Data can also be represented by using bar diagram or bar graph. In a bar graph, bars of uniform width are drawn horizontally or vertically. These bars are placed at equal distance from each other. The length of each bar gives the required information.

- The data in an unorganised form is called raw data. In order to draw meaningful inferences from a data, we need to organise the data systematically.

  We can organise a data in the following ways:

  o Frequency distribution table
  o Histogram
  o Pie chart

- **Construction of grouped frequency distribution table:**

  There are two ways to group the data to make frequency distribution table. These are as follows:

  Inclusive method (Discontinuous form):
  The classes can be defined in inclusive method as 1 - 10, 11 - 20, 21 - 30 and 31 - 40. Here, both limits are inclusive in each class.

  Exclusive method (Continuous form):
  In exclusive method, we take the class intervals as 0 – 10, 10 – 20, 20 – 30. The observations which are more than 0 but less than 10 will come under the group 0 – 10; the numbers which are more than 10 but less than 20 will come under the group 10 – 20 and so on. Here, the common observation will belong to the higher class, i.e. 10 will be included in the class interval 10 – 20 and similarly we follow this for the other observations also.
  For example, the ages of some residents of a particular locality are given as follows:
  7, 28, 30, 32, 18, 19, 37, 36, 14, 27, 12, 8, 17, 24, 22, 2, 21, 5, 21, 36, 38, 25, 10, 25, 9.
  Frequency distribution table can be drawn as follows:

  Inclusive method:

Class intervals	Tally marks	Frequency
1 – 10	|||||		6
11 – 20	|||||	5
21 – 30	||||| ||||	9
31 – 40	|||||	5

  Exclusive method:

Class intervals	Tally marks	Frequency
1 – 10	|||||	5
10 – 20	|||||		6
20 – 30	||||| |||	8
30 – 40	|||||		6

- **Few points to be remembered while choosing class intervals:**

1. Classes should not be overlapping and all values or observations should be covered in these classes.
2. The class size for all classes should be equal.
3. The number of class intervals is normally between 5 and 10.
4. Class marks and class limits should be taken as integers or simple fractions.