Statistics

- 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:

  - Frequency distribution table
  - Histogram
  - 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	𝍸 |	6
11 – 20	𝍸	5
21 – 30	𝍸 ||||	9
31 – 40	𝍸	5

Exclusive method:

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

1. Cumulative frequency for a class is the total of frequencies above or below it. Cumulative frequency (c.f.) can be of two types.

(i) Less than type

It is calculated to collect information about the frequency less than the upper limit of each class.

(ii) More than type

It is calculated to collect information about the frequency more than the lower limit of each class.

For example, for the given frequency distribution table, less than or more than type cumulative frequency can be calculated as follows:

Age (in years)	5 - 15	15 - 25	25 - 35	35 - 45
Number of persons	7	19	17	12

Age (in years) (Class)	**Number of persons (*f*)**	**Cumulative frequency (*c.f*) (less than type)**
5 - 15	7	7
15 - 25	19	7 + 19 = 26
25 - 35	17	26 + 17 = 43
35 - 45	12	43 + 12 = 55
Total	55	

Age (in years) (Class)	**Number of persons (*f*)**	**Cumulative frequency (*c.f*) (more than type)**
5 - 15	7	48 + 7 = 55
15 - 25	19	29 + 19 = 48
25 - 35	17	12 + 17 = 29
35 - 45	12	12
Total	55	

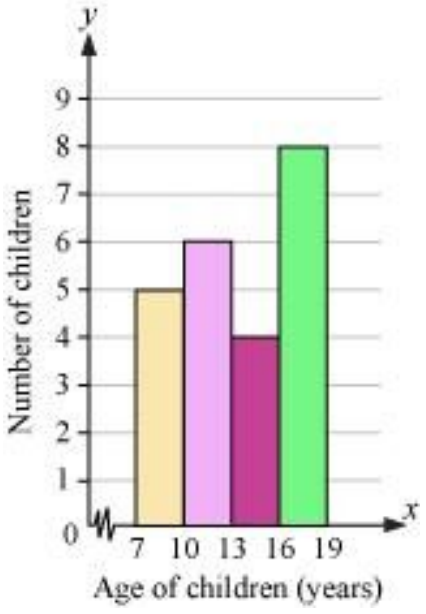
- **Histogram**

  A histogram is a bar graph that is used to represent grouped data. In a histogram, the class intervals are represented on the horizontal axis and the heights of the bars represent frequency. Also, there is no gap between the bars in a histogram.

Class interval	**Tally mark**	**Frequency**

(Age of children)		(Number of children)
7 – 10	𝍸	5
10 – 13	𝍸 |	6
13 – 16	𝍸	4
16 – 19	𝍸 𝍸	8

The above frequency distribution table can be displayed in a histogram as follows:



In a histogram, a broken line can be used along the horizontal axis to indicate that the numbers between 0 to7 are not included.

- **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.

  - 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.

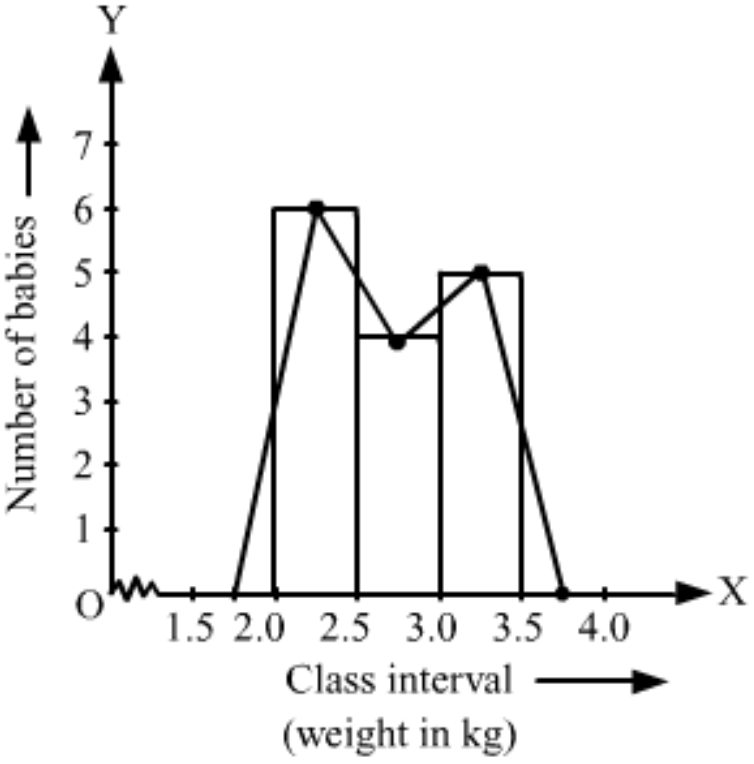
- One other way of drawing a frequency polygon is by plotting and joining the ordered pairs (of class marks and their corresponding frequencies) with the mid-points of the classes preceding and succeeding lowest and highest class respectively.

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 x1,x2,x3....xn is $\bar{x}$ then x1-$\bar{x}$+x2-$\bar{x}$+x3-$\bar{x}$+...+xn-$\bar{x}$=0.
2. If the mean of *n* observations x1,x2,x3....xn is $\bar{x}$ then the mean of x1+p, x2+p, x3+p, ..., xn+p is ($\bar{x} + p$).
3. If the mean of *n* observations x1,x2,x3....xn is $\bar{x}$ then the mean of x1-p, x2-p, x3-p, ..., xn-p is ($\bar{x} - p$).
4. If the mean of *n* observations x1,x2,x3....xn is $\bar{x}$ then the mean of px1, px2, px3, ..., pxn is $p\bar{x}$.
5. If the mean of *n* observations x1,x2,x3....xn is $\bar{x}$ then the mean of x1p, x2p, x3p, ..., xnp is $\bar{x}p$.

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

∴ Average runs scored = $\frac{426}{6} = 71$

- **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)

$$\therefore \text{Median} = \left(\frac{7+1}{2}\right)^{th} \text{observation}$$

$$= 4^{th} \text{observation}$$

$$= 30$$