Data Handling

Data is defined as a collection of numbers which give the required information. For example, marks scored by the students in a class, number of members in a family, number of books sold etc.

Data are of two types:

(i) **Primary data:** It is the data collected by the person directly for a specific purpose without referring any source. Primary data is collected through surveys, local sources etc.

(ii) **Secondary data:** It is the data collected through other sources like research organizations, financial institutions etc.

The original form of data is called **raw data**. But when the data is arranged in ascending or descending order, it is referred to as **array**.

For example, let us consider the following data.

Name of the student	Marks obtained (out of 100)
Manasi	81
Praveen	73
Pradeep	98
Kartik	61
Mamta	96
Vinod	83
Salma	69
Jyoti	83
Amardeep	67
Suraj	52

This data gives information about the marks obtained (out of 100) by 10 students.

By observing this data, we can say that Mamta obtained the highest marks and Suraj obtained the least marks among all the students.

We can also say that Jyoti and Vinod obtained the same marks.

- We arrange any data in tabular form using tally marks to obtain particular information in very little time.

For 1, we use the tally mark |

For 2, we use the tally mark ||

For 3, we use the tally mark |||

For 4, we use the tally mark ||||

For every 5, we use the tally mark 𝍸

Example: In order to understand the concept of tally marks, let us arrange the following data using tally marks. The given data represents the number of blood donors of different blood groups in a blood donation camp.

We can represent the given data using tally marks.

Blood group	Number of donors
O	28
A	19
B	16
AB	12

Solution: Using tally marks, the given data can be arranged as:

Blood group	Tally marks	Number of donors

O	𝍸 𝍸 𝍸 𝍸 𝍸 |||	28
A	𝍸 𝍸 𝍸 ||||	19
B	𝍸 𝍸 𝍸 |	16
AB	𝍸 𝍸 ||	12

This is known as a **tally chart**.

- **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 $x_1, x_2, x_3 \ldots x_n$ x1,x2,x3....xn is $\bar{x}$ $\bar{x}$ then $(x_1 - \bar{x}) + (x_2 - \bar{x}) + (x_3 - \bar{x}) + \ldots + (x_n - \bar{x}) = 0$ x1-$\bar{x}$+x2-$\bar{x}$+x3-$\bar{x}$+...+xn-$\bar{x}$=0.
2. If the mean of *n* observations $x_1, x_2, x_3 \ldots x_n$ x1,x2,x3....xn is $\bar{x}$ $\bar{x}$ then the mean of $(x_1 + p),\ (x_2 + p),\ (x_3 + p),\ \ldots,\ (x_n + p)$ x1+p, x2+p, x3+p, ..., xn+p is ($\bar{x}$ $\bar{x}$ + *p*).
3. If the mean of *n* observations $x_1, x_2, x_3 \ldots x_n$ x1,x2,x3....xn is $\bar{x}$ $\bar{x}$ then the mean of $(x_1 - p),\ (x_2 - p),\ (x_3 - p),\ \ldots,\ (x_n - p)$ x1-p, x2-p, x3-p, ..., xn-p is ($\bar{x}$ $\bar{x}$ − *p*).
4. If the mean of *n* observations $x_1, x_2, x_3 \ldots x_n$ x1,x2,x3....xn is $\bar{x}$ $\bar{x}$ then the mean of $px_1,\ px_2,\ px_3,\ \ldots,\ px_n$ px1, px2, px3, ..., pxn is *p*$\bar{x}$ $\bar{x}$.
5. If the mean of *n* observations $x_1, x_2, x_3 \ldots x_n$ x1,x2,x3....xn is $\bar{x}$ $\bar{x}$ then the

mean of $\frac{x_1}{p}, \frac{x_2}{p}, \frac{x_3}{p}, \ldots, \frac{x_n}{p}$ x1p, x2p, x3p, ..., xnp is $\frac{\bar{x}}{p}$ $\bar{x}p$.

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 = $\frac{426}{6} = 71$

- **Mode**
  1. The mode of a set of observations is the observation that occurs most often.
  2. Mode of a large data can be calculated by forming a tally marks table.

Example: What is the mode of data: 247, 346, 335, 247, 335, 346, 247, 335, 346, 351, 351, 346, 247, 247, 346, and 247?

Solution: A tally marks table for the given data is as follows.

Data	Tally marks	Frequency
247		6
335		3
346		5
351		2

Therefore, mode of the data is 247.

- 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.

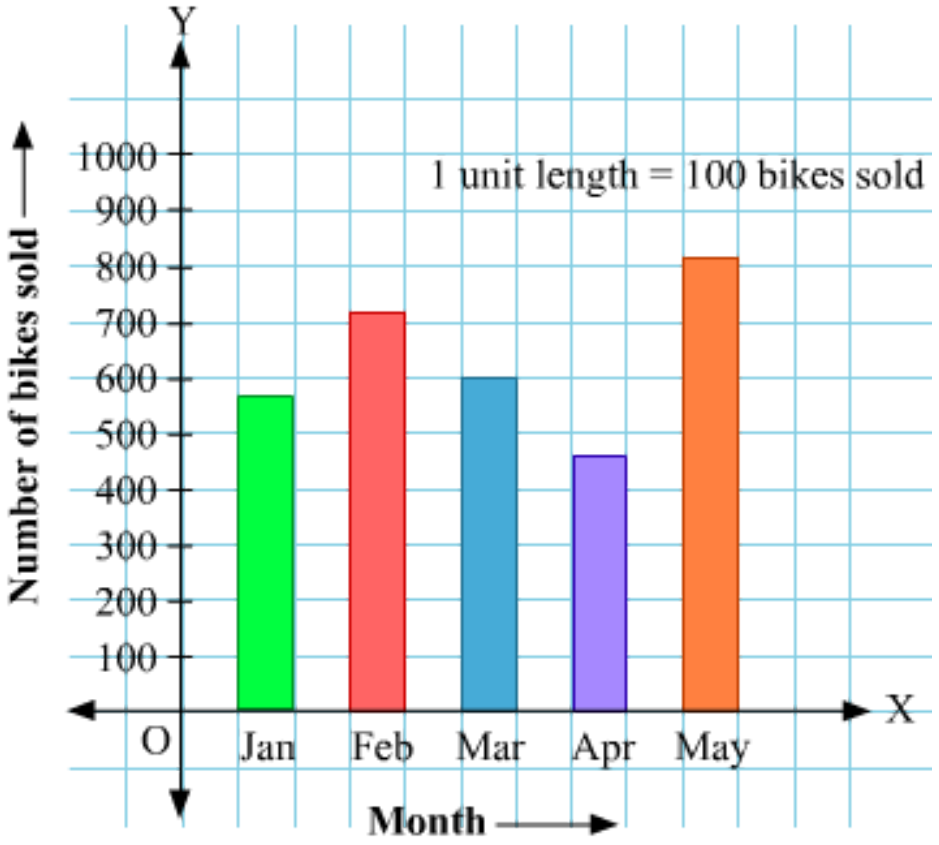
  Example:
  The given data represents the number of bikes sold by a retailer in the first five months of a year. Construct a bar graph of this data.

Month	Number of bikes sold
January	560
February	720
March	600
April	450
May	820

  Solution:
  To draw the bar graph for the given data, we proceed as follows:

  - Draw two perpendicular lines, one vertical and one horizontal
  - Mark the months along the horizontal line and mark the corresponding number of bikes along the vertical line.
  - Draw bars of same width and maintain uniform gaps between them.
  - Choose a suitable scale along the vertical line. Let 1 unit length = 100 bikes sold and mark the corresponding values.

  Hence, the bar graph of the given data can be drawn as:


Y
1000
900
800
700
600
500
400
300
200
100
1 unit length = 100 bikes sold
Number of bikes sold
X
O
Jan
Feb
Mar
Apr
May
Month