Chapter 3

Data Handling

Introduction to Data

A data is a collection of numbers, words, measurements, or just a description of things, gathered to give some information.

In our daily life, we see several kinds of data through newspapers, magazines, television, and other sources.

Example:

i) Your performance in annual examination.

ii) The number of storybooks read by Ethan and his sister during summer vacations.

iii) Marks scored by students of Class VII in the Mathematics Annual Examination.

Marks of 15 students of Class VII in Mathematics Annual Examination (out of 100):
78, 81, 57, 92, 88, 67, 94, 97, 82, 87, 75, 64, 96, 83, 86

Why do we collect data?

We first collect and record data, then it is presented in a way such that it gives meaningful information. We primarily collect data for the following reasons,

1) Analysis

Let us consider an example where we are using data for analysis.

The table given below is the data of the top 5 scorers in the FIFA World Cup 2018.

Scorer	Goals
Harry Kane	6
RomeluLukaku	4
Denis Cheryshev	4
Christiano Ronaldo	4
ArtemDzyuba	3

This data helps us to analyze the performance of the top 5 scorers in the FIFA World Cup 2018.

2) Comparison

Data is also useful for comparison of data sets.

The table given below shows the performance of Indian Cricketers.

Name	No. of Matches	Runs Scored	Highest Score	Average	Strike rate
M S Dhoni	341	10500	183	50.72	87.55
KedarJadhav	59	1174	120	43.48	102.53
Dinesh Karthik	91	1738	79	31.03	73.70

From the given table we can easily compare the performances of the cricketers. We can find which cricketer among these three have played maximum matches, whose strike rate is the highest?

3) Prediction

We use data for predictions, like weather forecasting where previous data of weather conditions are compared with the present data to predict future weather.

Collecting Data

The District Football Association has recorded the heights (in cm) ofthe football players of the district team. The heights of the playersare:152, 151, 148, 163, 159, 150, 143, 157, 147 and 167.

The above data can tell us many things, like the height of the tallest and the shortest player, but is this data sufficient if we want to find the age of the youngest player in the district football team?

The answer is no because for that we need to collect data regarding the age of all the players. Therefore, before collecting the data, we need to know what we would use it for.

Organisation of Data

The collection, recording, and presentation of data help us to draw inferences from them.

Now the District Football Association recorded the ages (in years) of the football players as well.

The ages (in years) of the players are:
13, 14, 14, 17, 16, 15, 16, 12, 15, 16.
The data in this form is called raw data.

Raw data is the data obtained in its original form.

Now, if we arrange the data in ascending or descending order, then we can interpret the data in a better way.

12, 13, 14, 14, 15, 15, 16, 16, 16, 17 (Ascending order)

The data arranged in an ascending or descending order is called an array or array data.

Now, we can easily tell that

- The youngest player in the team is 12 years old
- The oldest player is 17 years old.

The above data can also be arranged in a tabular form

Name	Ron	Ethan	Roger	Jacob	Ash	Shane	Charlie	Luca	Tom
Age (in years)	12	13	14	15	15	16	16	16	17

Arithmetic Mean

In our daily life, we come across statements like,

- The average age of children playing football in the district team is 16 years.
- Janet spends an average of 1 hour daily on her piano practice.

The first statement states that the average age of football players in the school team is 16 years.
Let's see the data, regarding the ages of the players, again.

Name	Ronald	Ethan	Roger	Jacob	Mark	Shane	Charlie	Lucas	Tony
Age (in years)	12	13	14	15	15	16	16	16	17

We see that most of the players on the football team are 16 years old, but some are more or less than 16 years. Average age does not mean that every player in the team is 16 years old.

In the second statement, the average time of 1 hour daily means that Janet usually practices for 1 hour, but on some days she practices for less than 1 hour and on other days she may practice longer.

Average is a number that represents or shows the central tendency (value) of a group of observations or data. Since average lies between the highest and the lowest value of the given data so, we say the average is a measure of the central tendency (value) of the group of data.

Example: Janice studies for 2 hours, 4 hours, 3 hours, and 2 hours respectively on four consecutive days. How many hours does he study daily on average?

Average time of Studies = $\frac{3+4+3+2}{4}$ hours = $\frac{12}{2}$ = 2 hours per day.

Therefore, we can say that Janice studies for 2 hours daily on an average.

Arithmetic Mean

Different types of data need different forms of representative or central value to describe it. One of these representative values is the "Arithmetic Mean".

The Arithmetic mean of a group of observations is defined as,

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

The mean of a group of observations is the value that is equally shared out among all the observations.

Example: A cricketer scores the following runs in eight innings: 58, 76, 40, 35, 46, 45, 0, 100. Find the mean score.

(REFERENCE: NCERT)

Runs scored by the cricketer in eight innings are 58, 76, 40, 35, 46, 45, 0 and 100.

Innings	1	2	3	4	5	6	7	8	Total Score
Runs Scored	58	76	40	35	46	45	0	100	400

Mean Score $= \frac{Total\ runs\ scored}{Total\ number\ of\ innings}$

$= \frac{58+76+40+35+46+45+0+100}{8} = \frac{400}{8} = 50$

We know that the mean of a group of observations is the value that is equally shared among all the observations.

Innings	1	2	3	4	5	6	7	8	Total Score
Runs Scored	50	50	50	50	50	50	50	50	400

We see that if we distribute the total score equally over each inning, then the batsman scored 50 runs in each inning.

Example: The following table shows the points of each player scored in four games:

(REFERENCE: NCERT)

Player	Game 1	Game 2	Game 3	Game 4
A	14	16	10	10
B	0	8	6	4
C	8	11	Did not play	13

Now answer the following questions:

i) Find the mean to determine A's average number of points scored per game.

ii) To find the mean number of points per game for C, would you divide the total points by 3 or by 4? Why?

iii) B played in all the four games. How would you find the mean?

iv) Who is the best performer?

i) Mean/average score of player A = $\frac{Total\ points\ scored}{Total\ number\ of\ games}$

$= \frac{14+16+10+10}{4} = \frac{50}{4} = 12.5$

ii) To find the mean number of points per game for C, we would divide his total score by 3 because player C played 3 games only.

iii) Mean score of player B = $\frac{Total\ points\ scored}{Total\ number\ of\ games}$

$= \frac{0+8+6+4}{4} = \frac{18}{4} = 4.5$

iv) The best performer will have the highest mean score. Now, we have already calculated the mean score of Player A and Player B. We need to find the mean score of Player C as well.

Mean score of player C = $\frac{Total\ points\ scored}{Total\ number\ of\ games}$

$= \frac{8+11+13}{3} = \frac{32}{3} = 10.6$

On comparing the mean score of Player A (12.5), Player B (4.5) and Player C (10.6) we get,

$12.5 > 10.6 > 4.5$

Therefore, Player A is the best performer as his mean score is the highest among the three players.

Range

The difference between the highest and the lowest observation is the range of the data.

Range = Highest Observation – Lowest Observation

Example: The marks (out of 100) obtained by a group of students in a Math test are 85, 76, 90, 85, 39, 48, 56, 95, 81 and 75. Find the:

(i) Highest and lowest marks obtained by the students.

(ii) Range of the marks obtained.

(iii) Mean marks obtained by the group.

Firstly, we arrange the marks obtained by the students in ascending order.

39, 48, 56, 75, 76, 81, 85, 85, 90 and 95

i) On arranging the marks in ascending order, we see that the highest mark obtained by the student is 95 and the lowest mark is 39.

Highest marks = 39

Lowest marks = 95

ii) Range is the difference between the highest and the lowest observation of the data.

Range = Highest Observation – Lowest Observation

Range = 95 – 39 = 56

iii) Mean marks = $\frac{Total\ marks\ obtained}{Number\ of\ students}$

$$= \frac{39+48+56+75+76+81+85+85+90+95}{10} = \frac{730}{10} = 73$$

Example: The heights of 10 girls of 8th class were measured in cm and the results are as follows: 135, 150, 139, 128, 151, 132, 146, 149, 143, and 141.

i) What is the height of the tallest girl in 8th class?

ii) What is the height of the shortest girl in 8th class?

iii) What is the range of the data?

iv) What is the mean height of the girls in 8th class?

v) How many girls have heights more than the mean height?

On arranging the heights of 10 girls in ascending order we get,

128, 132, 135, 139, 141, 143, 146, 149, 150 and 151

i) Height of the tallest girl is 151 cm

ii) Height of the shortest girl is 128 cm

iii) Range = Highest Observation – Lowest Observation

Range = 151 cm – 128 cm = 23 cm

iv)Mean height = $\frac{Sum\ of\ the\ heights\ of\ the\ girls}{Total\ number\ of\ girls}$

$$= \frac{128+132+135+139+141+143+146+149+150+151}{10} = \frac{1414}{10} = 141.4$$ cm

v) The heights of girls more than the mean height, 141.4 cm are 143 cm, 146 cm, 149 cm, 150 cm, and 151 cm.

Clearly, 5 girls have heights more than the mean height.

Mode

The owner of a bookshop has put up a stall in one of the Book Fair organized by the city mayor. He keeps a daily record of sales of all the books. The table given below shows the record of the weekly sale.

Name of the book	Matilda by Roald Dahl	Charlie and the Chocolate Factory by Roald Dahl	Pride and Prejudice by Jane Austen	Rebecca by Daphne Du Maurier	Great Expectations by Charles Dickens	Anna Karenina by Leo Tolstoy	Jane Eyre by Charlotte Bronte
Number of books sold	8	12	7	6	10	8	5

He saw the record and realized that Charlie and the Chocolate Factory and theGreat Expectations are more popular books among children and adults respectively. He decided to procure these two books, as their sale may increase during the coming weekend.

We see that the owner is concerned about the number of books sold and which books are sold the most during the week. This is another representative value for the data. This representative value is called the mode of the data.

The mode of a set of observations is the observation that occurs most often.

Consider the given set of numbers: 3, 3, 4, 5, 2, 2, 2, 2

We first arrange the given numbers in ascending order, keeping the numbers with the same value together.

2, 2, 2, 2, 3, 3, 4

Number	2	3	4
Frequency	4	2	1

Frequency is the number of times an observation occurs in the given data.

We see that 2 is repeated the maximum number of times, therefore 2 is the mode of the given data.

Example: Heights (in cm) of 30 children are given below:

165, 165, 163, 160, 163, 161, 162, 164, 163, 162, 164, 163, 160, 165, 164, 165, 166, 162, 163, 164, 166, 160, 166, 163, 162, 161, 160, 161, 163, 166

What is the mode of their heights?

We first put the data in tabular form:

Height (in cm)	Tally Bars	Number of children
160	||||	4
161	|||	3
162	||||	4
163	~~||||~~ | | |	7
164	||||	4
165	||||	4
166	||||	4
	Total	30

We see that 163 occurs the maximum number of times, so 163 cm is the mode of the given data.

Median

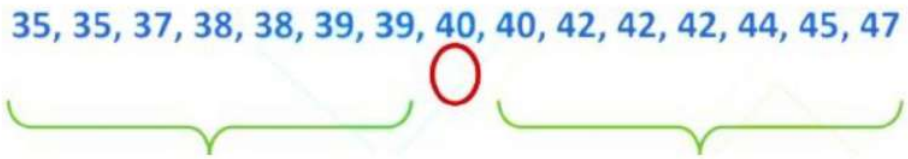
A class teacher is conducting a math quiz in the class. Now, she wants to divide students of the class into two groups, so that both the groups have an equal number of students. How would she do it?

She decided to form two groups based on the weights of the students. One group has students with weights lesser than a particular weight and the other group has students with weights more than a particular weight.

The weights (in kg) of the students are as follows:

35, 38, 37, 35, 39, 40, 42, 38, 39, 40, 42, 45, 47, 42, 44

She first arranged the weights of the students in ascending order.



She saw that the middle value of the given data, 40 kg is the value that divides the data into two equal groups. This middle value is called the Median.

Median is the value that lies in the middle of the data, arranged in an ascending or descending order.

So, the class teacher divided the students into two groups of 7 students each.

Are you wondering what happened to the student whose weight is the middle value? Well, the teacher decided to make him the scorekeeper.

1) If the number of observations(n) in the given data is odd then,

Median = Value of $\left(\frac{n+1}{2}\right)^{th}$ observation

2) If the number of observations(n) in the given data is even then,

Median = $\frac{\text{Value of } \left(\frac{n}{2}\right)^{th} \text{ observation} + \text{Value of } \left(\frac{n}{2}+1\right)^{th} \text{ observation}}{2}$

Example: The runs scored in a colony cricket match by 11 boys is as follows: 6, 15, 120, 50, 100, 80, 10, 15, 8, 10, 15

Find the mean, mode, and median of this data. Are the three same?

We first arrange the given scores in ascending order.

6, 8, 10, 10, 15, 15, 15, 50, 80, 100, 120

i) Mean score = $\frac{Sum\ of\ the\ scores\ of\ players}{Total\ number\ of\ players}$

$= \frac{6+8+10+10+15+15+15+50+80+100+120}{11} = \frac{429}{11}$

Mean Sore = 39

ii) The given scores are 6, 8, 10, 10, 15, 15, 15, 50, 80, 100, 120.

We see that 15 occurs the maximum number of times. Therefore, the mode is 15

iii) The scores are: 6, 8, 10, 10, 15, 15, 15, 50, 80, 100, 120

The number of observations is 11, an odd number.

So, Median = $\left(\frac{n+1}{2}\right)^{th}$ Value of observation

Median = Value of $\left(\frac{11+1}{2}\right)^{th} = \left(\frac{12}{2}\right)^{th} = 6^{th}$ observation

We see that the value of the 6^{th} observation in the given data is 15.

Therefore, 15 is the median of the given data.

Example: The weights (in kg) of 15 boys in a class are:

38, 42, 35, 37, 45, 50, 32, 43, 43, 40, 36, 38, 43, 38, 47

i) Find the mode and median of the weights.

ii) Is there more than one mode?

We first arrange the data in ascending order.

32, 35, 36, 37, 38, 38, 38, 40, 42, 43, 43, 43, 45, 47, 50

i) We see that both 38 and 43 are repeated the maximum number of times. Therefore, 38 and 43 are the modes of the given data.

Number of observations in the given data is 15, an odd number.

So, Median = Value of $\left(\frac{n+1}{2}\right)^{th}$ observation

Median = Value of $\left(\frac{15+1}{2}\right)^{th} = \left(\frac{16}{2}\right)^{th} = 8^{th}$ observation

We see that the value of the 8th observation in the given data is 40.

Therefore, 40 is the median of the given data.

ii) Yes, there are more than two modes. Both 38 and 43 occur three times in the given data. Therefore, 38 and 43 are the modes of the given data.

Bar Graph

A bar graph is a pictorial representation of numerical data using bars of uniform width drawn horizontally or vertically with equal spacing between them. The length of the bar depends upon the number it represents.

A bar graph is an effective way of presenting visual information. We see bar graphs in newspapers, magazines comparing sets of data between different groups.

Students of Class VII were asked to name their favourite ice-cream flavour as the school committee had decided to give ice – creams to students on Children's Day. The following bar graph shows the preferences of the students.

Ice Cream Flavour	Vanilla	Chocolate	Strawberry	Mango
Number of Students	5	14	10	6

We see that the bar representing the number of students who like chocolate is the tallest.

So, students' most preferred ice cream is chocolate.

Constructions of Bar Graphs

Example: Numbers of students in six different classes are given below.
Represent the data on a bar graph.

Step 1: Draw two perpendicular lines. Mark them as OX and OY.
These lines are called the x −axis and y −axis respectively.

Step 2: Along OX, write classes at points taken at uniform gaps and along OY, write the number of children.

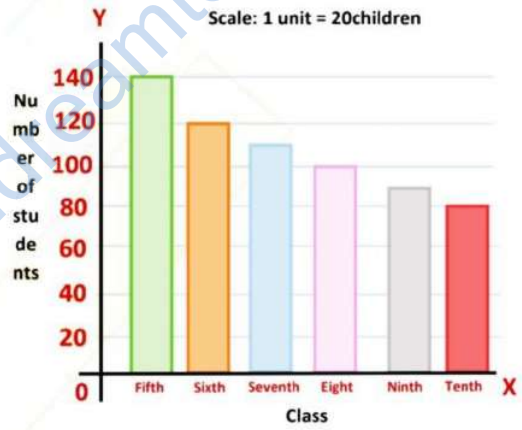
Class	Fifth	Sixth	Seventh	Eight	Ninth	Tenth
Number of students	140	120	110	100	90	80

Step 3: Choose a suitable scale on OY. We will choose a scale as

1 unit = 20 children

Step 4: Calculate the heights of the bar according to the scale chosen.

Class	Number of students	Height of the bar (1 division = 20 students)
Fifth	140	$\frac{140}{20} = 7$
Sixth	120	$\frac{120}{20} = 6$
Seventh	110	$\frac{110}{20} = 5.5$
Eight	100	$\frac{100}{20} = 5$
Ninth	90	$\frac{90}{20} = 4.5$
Tenth	80	$\frac{80}{20} = 4$



Example: Use the bar graph to answer the following questions.

i) Which is the most popular pet?

ii) How many children have a dog as a pet?

i) We see that the bar representing the number of students who have a cat as a pet is the tallest; therefore cat is the most popular pet.

ii) We know,

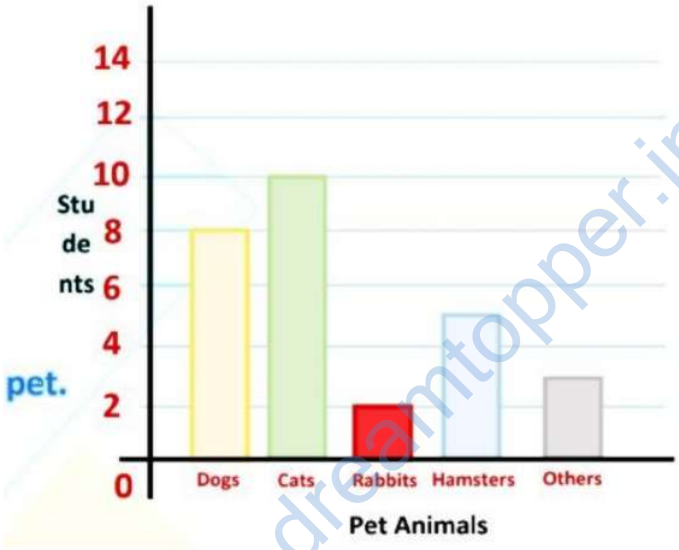
1 unit = 2 students

Height of the bar = 8 units

Number of Students

= 2× 8 = 16 students

So, 16 students have a dog as a pet.



Double Bar Graph

A double bar graph is a graphical display of information using two bars beside each other at various heights.

With the help of a double bar graph, we can compare two data groups.

The following double bar graph displays the number of Mathematics and Science books bought by a school library from the year 2015 – 2019.

Year	2015	2016	2017	2018	2019
Mathematics	30	35	25	40	50
Science	25	35	40	20	25

With the help of a double bar graph, we can easily compare the number of Mathematics and Science books bought in the last 5 years.

Construction of Double Bar Graph

Example: Consider this data collected from a survey of a colony.

Favourite Sport	Cricket	Basket Ball	Swimming	Hockey	Athletics
Watching	1240	470	510	430	250
Participating	620	320	320	250	105

i) Draw a double bar graph choosing an appropriate scale. What do you infer from the bar graph?

ii) Which sport is the most popular?

iii) Which is more preferred, watching or participating in sports?

(REFERENCE: NCERT)

Step 1: Draw two perpendicular lines. Mark them as OX and OY. These lines are called the x − axis and y − axis respectively.

Step 2: Along OX, write the name of the sports at points taken at uniform gaps and along OY, write the number of people.

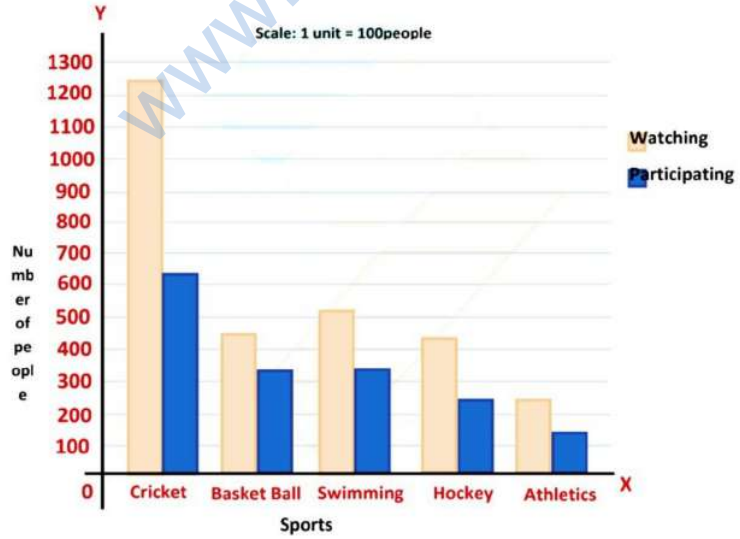
Sports	Watching		Participating	
	Number of People	Height of the bar (1 division = 100 children)	Number of People	Height of the bar (1 division = 100 children)
Cricket	1240	$\frac{1240}{100} = 12.4$	620	$\frac{620}{100} = 6.2$
Basket Ball	470	$\frac{470}{100} = 4.7$	320	$\frac{320}{100} = 3.2$
Swimming	510	$\frac{510}{100} = 5.1$	320	$\frac{320}{100} = 3.2$
Hockey	430	$\frac{430}{100} = 4.3$	250	$\frac{250}{100} = 2.5$
Athletics	250	$\frac{250}{100} = 2.5$	105	$\frac{105}{100} = 1.5$

Step 3: Choose a suitable scale on OY. We choose a scale as

1 unit = 100 people

Step 4: Calculate the heights of the bar according to the scale chosen.

Step 5: On the x – axis, draw pairs of bars of equal width and heights (calculated in Step 4)



This double bar graph displays the number of people who like watching and participating in different sports.

ii) The bar graph representing the number of people who like watching and participating in cricket is the tallest. So, cricket is the most popular sport.

iii) The bars representing watching sports are longer as compared to the bar representing participating in sports. Therefore, we can conclude that watching sports is preferred than participating in it.

Chance and Probability

The Chance of happening of an event may be described as one of the following:

We often hear these statements like:

- It will probably rain in the afternoon.
- Indian Cricket team will probably win the ICC World Cup series this year.

In all of the above statements, there is uncertainty about the occurrence of the event. The word 'probability' means there is uncertainty about the happening of the event.

Terms related to Probability

Experiment:

We need Tails on two coins when we throw them simultaneously.

Now, on tossing two coins simultaneously the possible outcomes are {H, H}, {H, T}, {T, H}, {T, T}.

We may not get Tails on both the coins in our first attempt. So, we repeat this process till we get Tails on both the coins.

Toss	1	2	3	4	5	6	7
Result	{H, T}	{T, H}	{H, H}	{H, T}	{H, H}	{H, T}	{T, T}

So, we finally got Tails on both the coins in our seventh attempt.

This whole process of tossing the coins simultaneously to get Tails on both the coins is called an experiment.

An experiment is an operation that produces some well-defined outcomes. E.g. tossing a coin, rolling a dice.

Each toss of the coins is called a trial.

A trial is an action that results in one or several outcomes.

In this experiment, we knew all the possible outcomes but we could not predict the exact outcome in advance. Such an experiment is called a random experiment.

An experiment in which all the possible outcomes are known and the exact outcomes cannot be predicted in advance is called a random experiment.

Example: Suppose a jar contains 7 green and 6 blue marbles. If a marble is drawn at random from the jar without seeing, can you tell what the probability of getting a green marble is?

Total number of marbles in the jar = 13

Number of green marbles = 7

Probability of getting a green marble = $\frac{7}{13}$