

Chapter – 14

Statistics

Introduction to Statistics

Introduction

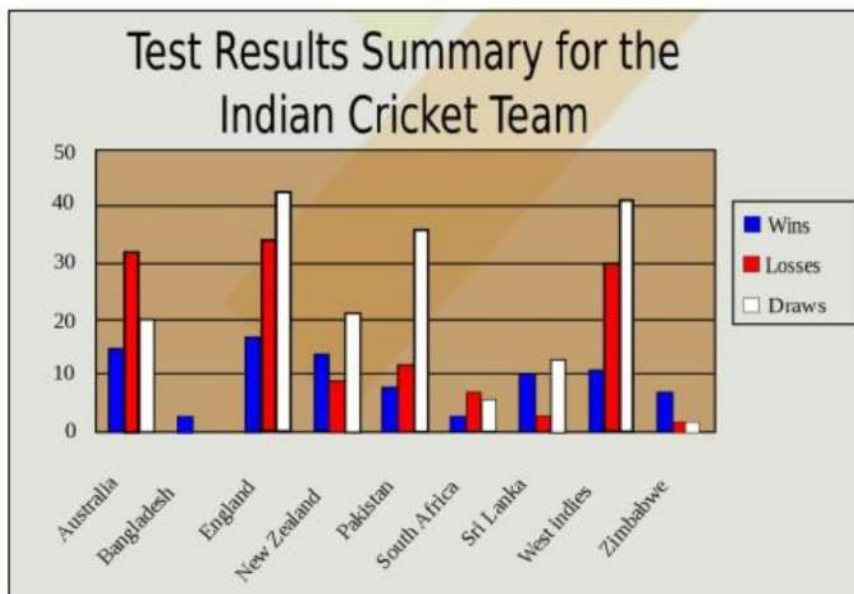
Every day we come across a lot of information in the form of facts, numerical figures, tables, graphs, etc. as shown figures below. For example:

- Runs scored by a team,
- Profits made by a company,
- Temperatures recorded in a day of a city,
- Expenditures in various sectors by government,
- The weather forecast, election results, and so on.

These facts or figures, which are numerical or otherwise, collected with a definite purpose are called data.

Therefore, Data is a collection of facts, such as numbers, words, measurements, observations, etc.

Suppose, the following image shows the performance of the Indian Cricket team in tests in the last 2 years against major test playing nations.



Types of data based on the collection of facts

Qualitative data: It is descriptive data. For examples:

- Rajan is thin.
- Suman can run fast.
- The cake is orange in color.
- She has black hair.
- He is tall.

Quantitative data: It is numerical information. For examples:

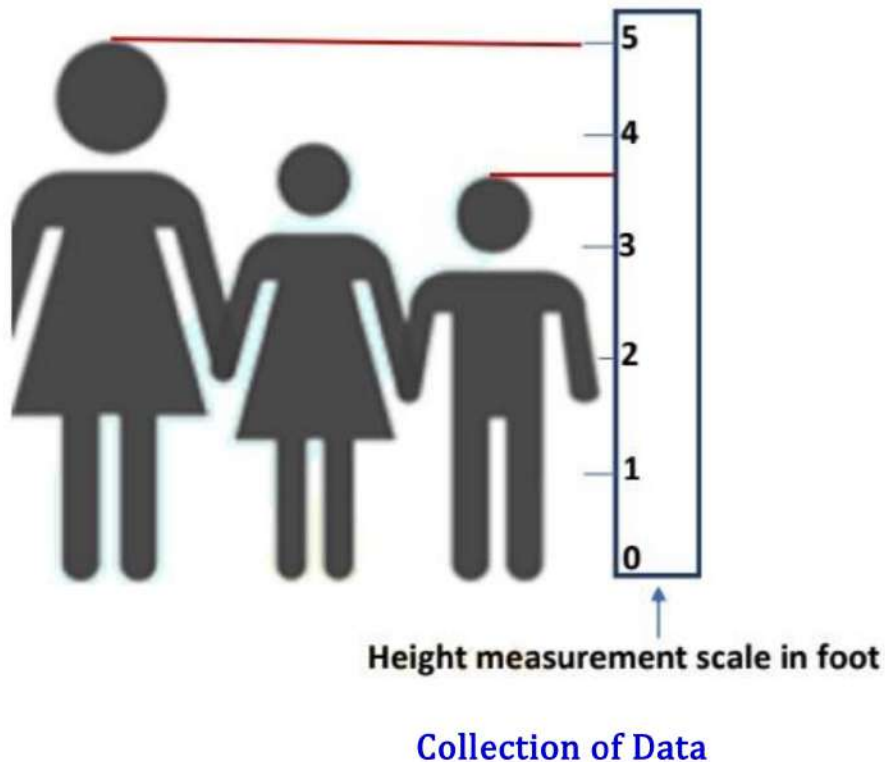
- I updated my phone 6 times in a quarter
- 83 people downloaded the latest mobile application
- She has 10 holidays in this year
- 500 people attended the seminar
- 54% of people prefer shopping online instead of going to the mall.

Types of Quantitative data

Discrete data: It has a particular fixed numerical value that can be counted. For example, in the image below, there are 12 students in the class. Here, the number of students in the class is a discrete data.



Continuous data: It does not have a fixed value but a range of data. For example, in the figure below, the height of the 3 persons lie between 3 feet to 5 feet.



Collection of Data

Our world is becoming more and more information oriented. Every part of our life utilizes information in one form or another. So, it becomes essential for us to know how to extract meaningful information from such data. The extraction of meaningful information is studied in a branch of mathematics called Statistics.

This involves the study of the collection, analysis, interpretation, presentation, and organization of data. In other words, it is a mathematical discipline to collect, summarize data.



Types of data on the basis of the collection of data

Primary data: It is the data that is collected by a researcher from first-hand sources, using methods like surveys, interviews, or experiments. For example, The following data is collected by a student for his/her thesis for the research project.

- Height of 10 students in your class.
- Number of absentees in each day in your class for a month.
- Number of members in the families of your classmates.
- Height of 10 plants in or around your home.

Secondary data: It is the data that has already been collected by someone, and then it is updated, tailored or modified for a specific purpose.

For example, in a school, the class-teachers of respective sections record attendance on a daily basis.

This data recorded by the class- teacher is an example of primary data. On a given day, the principal of the school asks for the attendance of all students of each section, to collate the total number of students present in the school on a given day.

This data collected by the school principal is an example of secondary data.

Presentation of data

Presentation of data

After the collection of data, we need to arrange the data such that it is easily understood, meaningful and serves the required purpose.

Such an arrangement is called the presentation of data. The raw data can be arranged in any of the following ways:

- Serial order or alphabetical order
- Ascending order
- Descending order
- Frequency Distribution

When the raw data is arranged in ascending or descending order, then the data is called an array or arrayed data.

Suppose that the marks obtained by 10 students of class 9th in a mathematics test, out of 50 marks according to their roll numbers

be: 39, 45, 33, 19, 21, 41, 21, 19, 40, 41.

The data in this form are called raw data or ungrouped data. The above raw data can be arranged according to in serial order (roll number) as follows:

Roll Number	Marks
1	39
2	45
3	33
4	19
5	21
6	41
7	21
8	19
9	40
10	41

Suppose, we want to find out who scored maximum or minimum marks in the test. The data in the given form does not give us a clear

understanding of the performance of the students.

If we arrange the marks scored in ascending or descending order, it gives us a better understanding of the given data. Also, we can easily identify the minimum and

maximum values in the data.

In ascending order, the data looks as follows:

19, 19, 21, 21, 33, 39, 40, 41, 41, 45. \Rightarrow Array or arrayed data.

In descending order, the data looks as follows:

45, 41, 41, 40, 39, 33, 21, 21, 19, 19. \Rightarrow Array or arrayed data.

From the above array or arrayed data, we can easily identify that the minimum marks are 19 and maximum marks are 45.

Frequency distribution

Suppose, 40 students appeared for a mathematics test. There is a possibility that the score of two or more students is the same. To find out how many students scored the same marks, we use a frequency distribution. If there is more than one instance like this, then finding that data may be time-consuming.

So, when the number of observations is large, then arranging the data in serial order or ascending order or descending order can be quite time consuming and also it does not tell us much except the minimum and maximum of the given data.

To minimize this effort and easily understand the data, we tabulate the data in a frequency distribution table.

Frequency distributions are of two types: Ungrouped frequency distribution table and Grouped frequency distribution table.

Let us consider a large data like the marks obtained (out of 100 marks) by 40 students of class 9th of a school.

50, 60, 70, 21, 19, 33, 39, 21, 92, 88, 80, 70, 72, 19, 40,
41, 92, 50, 50, 56, 60, 70, 60, 60, 88, 41, 45, 92, 88, 95,
70, 40, 39, 33, 19, 21, 41, 45, 70, 80.

If we arrange them in ascending order, it gives us a slightly better picture.

19, 19, 19, 21, 21, 21, 33, 33, 39, 39, 40, 40, 41, 41,
41, 45, 45, 50, 50, 50, 56, 60, 60, 60, 60, 70, 70, 70, 70,
70, 72, 80, 80, 88, 88, 88, 92, 92, 92, 95.

But, in the data arranged above, we cannot easily find how many students scored 41 marks or 60 marks. Again, we have to count that.

To make data easily understandable and clear, we can tabulate data of 40 students as shown below.

Marks	Tally marks	Number of students
19	III	3
21	III	3
33	II	2
39	II	2
40	II	2
41	III	3
45	II	2
50	III	3
56	II	1
60	IIII	4
70	IIII	5
72	I	1
80	II	2
88	III	3
92	III	3
95	I	1

Marks are called variates and the number of students who secured a particular number of marks is called the frequency of the variate.

The number of times an observation occurs in the given data is called the frequency of the observation.

This way of presentation of data is known as frequency distribution or ungrouped frequency distribution table.

In the above table, we can observe how many students scored the same marks but we have recorded 19 such observations. In the even bigger data than this, we may have to draw much bigger tables which makes our work cumbersome and time-consuming. To overcome this limitation, we can represent that data in classes or groups as shown below:

Marks	No of students (frequency)
11-20	3
21-30	3
31-40	6
41-50	8
51-60	5
61-70	5
71-80	3
81-90	3
91-100	4

In the table above, we have grouped the marks obtained by students in groups, which are called the classes or class intervals and their size is called the class-size or class width. Class 11-20 means the marks obtained between 11 and 20 including both 11 and 20. The number of observations falling in a particular class is called the frequency of that class or class frequency.

Class size of the above data is 9. We find it by finding the difference between the upper limit and the lower limit of the class interval.

For example, for the class interval 11-20, the smaller value 11 is the lower limit and the greater value 20 is the upper limit.

Now, the class-size or class width for the above data = $(20 - 11) = (30 - 21) = (40 - 31) = (50 - 41) = (60 - 51) = (70 - 61) = (80 - 71) = (90 - 81) = (100 - 91) = 9$.

This type of presentation of data is called a grouped frequency distribution table. Presenting data in this form simplifies and condenses data and enables us to observe certain important features at a glance.

We can prepare Grouped frequency distributions by two methods:

I. Inclusive Method

II. Exclusive Method

Inclusive Method: In this method, the classes are so formed that the upper limit of a class is included in that class. For example: In the

class 11-20 of marks obtained by students, a student who has obtained 20 marks is included in this class.

Grouped frequency distribution table:1 is shown below which is arranged by an inclusive method.

Marks	No of students (Frequency)
11-20	3
21-30	3
31-40	6
41-50	8
51-60	5
61-70	5
71-80	3
81-90	3
91-100	4

Table:1

Suppose, two new students are admitted in the class 9th whose marks are 20.5 and 30.5, but we cannot add them in the class

intervals viz. 11-20 and 21-30 as these values are not included in any of these class intervals. To include marks 20.5 and 30.5, we use

Exclusive Method for preparing grouped frequency distribution. So, to include marks 20.5 and 30.5, we need intervals such that the

upper limit of a class interval should be the same as the lower limit of the next class interval.

Exclusive Method: The class intervals are formed such that the upper limit of a class interval is the same as the lower limit of the next class interval. This method is called the exclusive method of classification.

Grouped frequency distribution table:2 is shown below which is arranged by an exclusive method.

Marks	No of students (frequency)
10-20	4
20-30	4
30-40	4
40-50	7
50-60	4
60-70	4
70-80	6
80-90	5
90-100	4

In this method, the upper limit of a class is not included in the class.

For example, if a student scores 20 marks, then it is included in the class 20-30 but not in the class 10-20. So, any observation which is

common to two class intervals, then it shall be considered in the higher-class interval.

Example: Given below are the ages of 25 students of class 9th in a school. Prepare a ungrouped frequency distribution and grouped

frequency distribution table.

15, 16, 16, 17, 17, 16, 15, 15, 16, 16, 17, 15, 16, 16, 14, 16, 15, 14, 16, 15, 14, 15, 16, 16, 15, 14, 15.

Solution: In the given data the observations are only 14, 15, 16 and 17. These ages are repeated multiple times. So, 14, 15, 16 and 17 are

variates of data.

Frequency distribution of the ages of 25 students is given below.

Age	Tally marks	Frequency
14	IIII	4
15	HH III	8
16	HH HH	10
17	III	3
Total		25

For a grouped frequency distribution table, we decide class interval according to own convenience.

Grouped frequency distribution table including exclusive method of ages of 25 students are given below.

Class interval	Tally marks	Frequency
13-15	IIII	4
15-17	HH HH HH III	18
17-19	III	3
Total		25

Graphical Representation of Data

Graphical Representation of Data

A graphical representation is the visual display of data and its statistical results. It is more often and effective than presenting

data in tabular form. Bar graphs, Histograms, and frequency polygons are different types of graphical representation, which

depends on the nature of the data and the nature of statical results.

Bar Graphs: Bar graphs are the bars of uniform width that can be drawn horizontally or vertically with equal spacing between them and then the length of each bar represents the given number. Such a method of representing data is called a bar diagram or a bar graph.

For a clear representation of categorical data or any ungrouped discrete frequency observations, we generally use the bar graphs.

Example 1: Considering the modes of transport of 30 students of class 9th is given below:

Student	Mode	Student	Mode	Student	Mode
Aman	Car	Tabish	Metro	Karan	Car
Iram	Scooty	Nauman	Metro	Raj	Bike
Rajan	Metro	Sana	Car	Sameer	Metro
Richa	Metro	Salman	Bike	Manish	Bus
Snehal	Bus	Sumaiya	Metro	Arif	Bike
Yusra	Metro	Diksha	Bus	Amir	Metro
Asna	Car	Ruby	Car	Asif	Car
Farha	Scooty	Arman	Bike	Kuldip	Bike
Bushra	Scooty	Rakesh	Metro	Ramesh	Metro
Shahrukh	Metro	Raman	Bus	Mukesh	Bus

In order to draw the bar graph for the data above, we prepare the frequency table as given below.

Mode of transport	Number of students
Car	6
Bike	5
Scooty	3
Metro	11
Bus	5

Now, we can represent this data using a bar graph, by following the steps as shown below:

- First, we draw two axes viz. x-axis and y-axis. Then, we decide what each axis of the graph represents. By convention,

the variates being measured goes on the horizontal (x-axis) and the frequency goes on the vertical (y-axis).

- Next, decide on a numeric scale for the frequency axis. This axis represents the frequency in each category by its height. It

must start at zero and include the largest frequency.

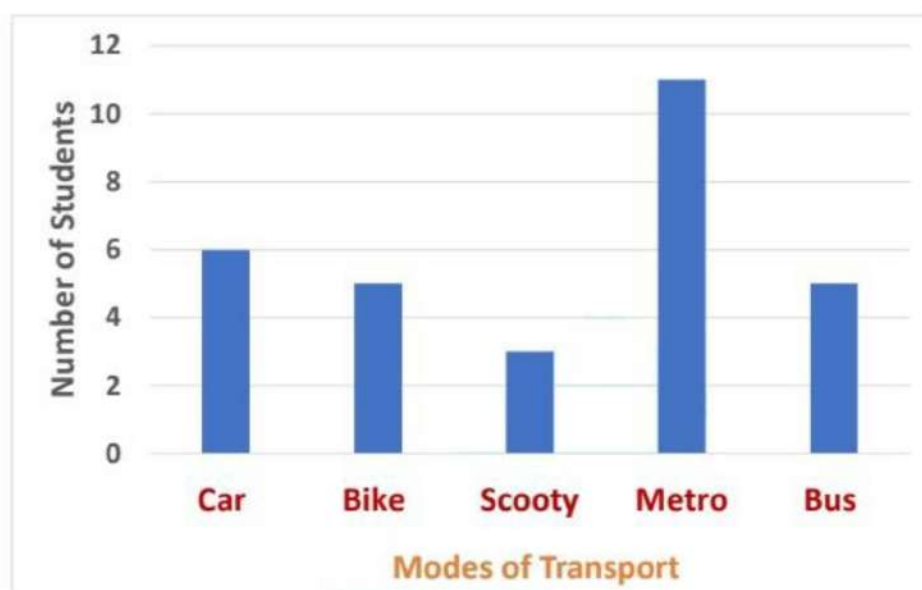
- Having decided on a range for the frequency axis we need to decide on a suitable number scale to label this axis. This

should have sensible values, for example, 0, 1, 2, . . . , or 0, 10, 20 . . . , or other such values as to make sense given the data.

- Draw the axes and label them appropriately.
- Draw a bar for each category. When drawing the bars it is essential to ensure the following:

- the width of each bar is the same.

- the bars are separated from each other by equally sized gaps.



Using this bar graph, we can easily identify the most popular mode of transport is the metro. Bar graphs provide a simple method of quickly spotting patterns within a discrete data set.

Histograms

Histogram was first introduced by Karl Pearson in 1891. Bar charts have their limitations; like they cannot be used to represent

continuous data. When dealing with continuous random variables different kinds of graphs are used. This type of graph is called a histogram.

At first sight, a histogram looks similar to bar charts. However, there are two critical differences:

- The horizontal (x-axis) is a continuous scale. As a result of this, there are no gaps between the bars (unless there are no observations within a class interval).
- The height of the rectangle is only proportional to the frequency of the class if the class intervals are all equal. With histograms, it is

the area of the rectangle that is proportional to their frequency.

Example 2: Consider the weights of 20 students of a class 9th as given below:

Students	Weights	Students	Weights
Aman	40	Tabish	53
Iram	42	Nauman	47
Rajan	55	Sana	46
Richa	41	Salman	60
Snehal	58	Sumaiya	59
Yusra	64	Diksha	42
Asna	57	Ruby	52
Farha	62	Arman	53
Bushra	61	Rakesh	43
Shahrukh	57	Raman	46

Now, arranging the data in ascending order.

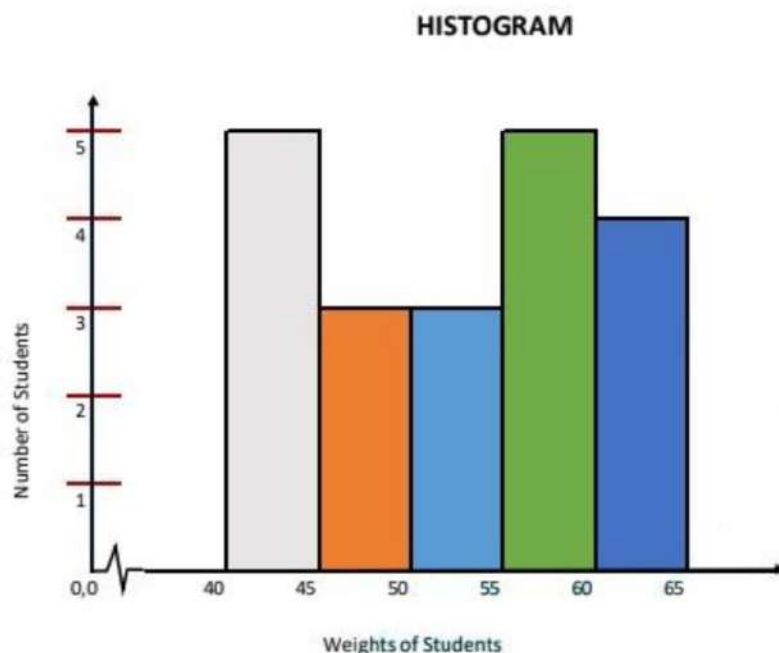
40, 41, 42, 42, 43, 46, 46, 47, 52, 53, 53, 55, 57, 57, 58, 59, 60, 61, 62, 64.

In order to draw the histogram for the data above, we prepare the frequency table as given below.

Weights (in kg)	Number of students
40-45	5
45-50	3
50-55	3
55-60	5
60-65	4

We can represent this information using histogram, by following steps as shown below:

- Find the maximum frequency and draw the vertical (y-axis) from zero to this value.
- The range of the horizontal (x-axis) needs to include a full range of the class intervals from the frequency table.
- Draw a bar for each group in your frequency table. These should be the same width and touch each other (unless there are no data in one particular class).



Frequency Polygon

It is a natural extension of the histogram. In frequency polygon rather than drawing bars, each class is represented by one point

and these are joined together by straight lines. We draw frequency polygons in a similar way of drawing a histogram.

Example 3: Consider the weights of 20 students of a class 9th as given below:

Students	Weights	Students	Weights
Aman	40	Tabish	53
Iram	42	Nauman	47
Rajan	55	Sana	46
Richa	41	Salman	60
Snehal	58	Sumaiya	59
Yusra	64	Diksha	42
Asna	57	Ruby	52
Farha	62	Arman	53
Bushra	61	Rakesh	43
Shahrukh	57	Raman	46

Now, arranging the data in ascending order.

40, 41, 42, 42, 43, 46, 46, 47, 52, 53, 53, 55, 57, 57, 58, 59, 60, 61, 62, 64.

In order to draw the frequency polygon for the data above, we prepare the frequency table as given below.

Weights (in kg)	Number of students
40-45	5
45-50	3
50-55	3
55-60	5
60-65	4

We can then present this information as a frequency polygon, by following the process of the steps shown below:

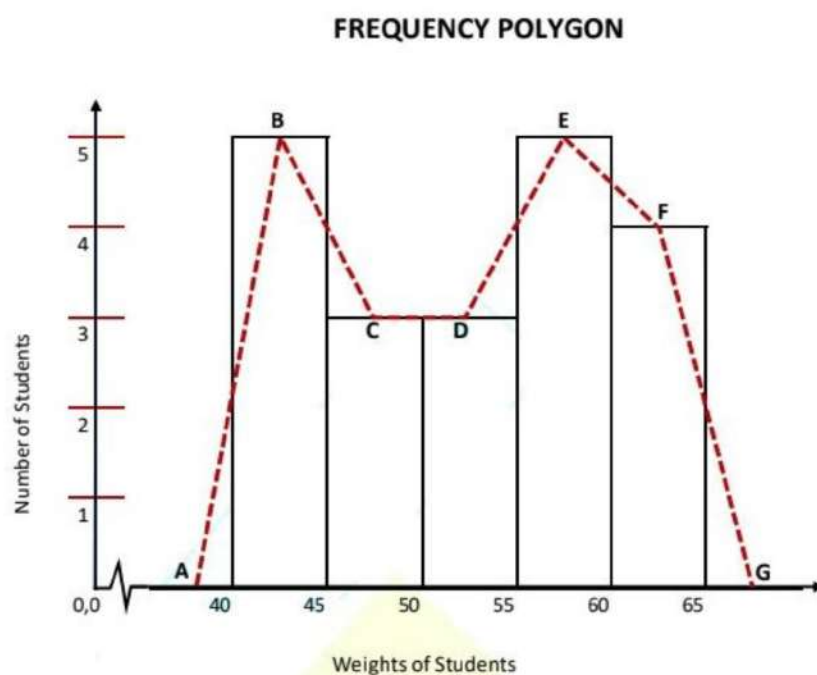
- Prepare a frequency table.
- Find the maximum frequency and draw the vertical (y-axis) from zero to this value.

- The range of the horizontal (x-axis) needs to include all class intervals from the frequency table.

- Draw bars for each class interval in the frequency table. These bars should be of the same width and are adjacent to each other (unless there are no data in one particular

class).

- Connect the midpoints of the top side of each bar by a dotted line as shown below.



Frequency polygons can also be drawn independently without drawing histograms. For this, we require the mid-points of the

class-interval. These mid-points of the class intervals are called class marks.

$$\text{Class mark} = \frac{\text{Upperlimit} + \text{Lowerlimit}}{2}.$$

Example 4: Consider the weights of 20 students of a class 9th as given below:

Students	Weights	Students	Weights
Aman	40	Tabish	47
Iram	42	Nauman	47
Rajan	44	Sana	46
Richa	41	Salman	41
Snehal	44	Sumaiya	43
Yusra	48	Diksha	42
Asna	44	Ruby	40
Farha	42	Arman	43
Bushra	43	Rakesh	43
Shahrukh	43	Raman	45

In ascending order, the data looks as follows:

40, 40, 41, 41, 42, 42, 42, 43, 43, 43, 43, 43, 44, 44, 44, 45, 46, 47, 47, 48

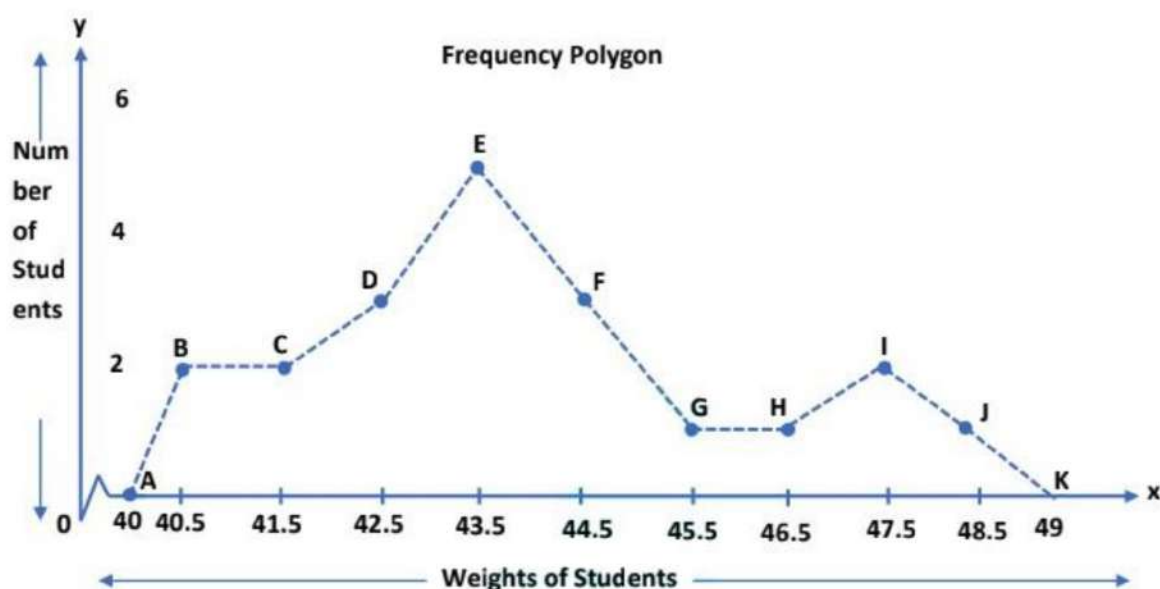
In order to draw the frequency polygon for the data above, we prepare the frequency table as given below.

Weights (in kg)	Class-marks	Number of students
40-41	40.5	2
41-42	41.5	2
42-43	42.5	3
43-44	43.5	5
44-45	44.5	3
45-46	45.5	1
46-47	46.5	1
47-48	47.5	2
48-49	48.5	1

We can then present this information as a frequency polygon, by following the process of the steps shown below:

- Prepare a frequency table.

- Find the maximum frequency and draw the vertical (y-axis) from zero to this value.
- The value of the horizontal (x-axis) needs to include all value of the class-marks from the frequency table.
- Connect the class mark of each bar by a dotted line as shown below.



Measures of Central Tendency

A measure of central tendency represents the center point of data. These measures indicate where most values in a distribution lie and are also referred to as the central location of a distribution.

There are three main measures of central tendency: the mean, the median, and the mode.

Mean (Average): It is calculated by dividing the sum of all observations in data by the number of observations. So, if we have n observations in a data set and they have observations x_1, x_2, \dots, x_n , the sample mean, usually denoted by \bar{x} (read as x bar), is:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

$$\text{Mean} = \frac{\text{Sum of all the observations}}{\text{Total number of observations}}$$

This formula is usually written in a slightly different manner using the Greek capital letter, Σ , read as "sigma", which means sum of:

$$\bar{x} = \frac{\sum x}{n} \text{ . Where } x = x_1 + x_2 + \dots + x_n.$$

Example 1: Find the mean of the marks obtained by 20 students of class 9th of a school :

20, 30, 30, 10, 40, 45, 30, 20, 25, 45, 10, 25, 35, 45, 40, 20, 30, 25, 20, 10.

Solution: Suppose that $x_1 = 20, x_2 = 30, x_3 = 30, x_4 = 10, x_5 = 40, x_6 = 45, x_7 = 30, x_8 = 20, x_9 = 25, x_{10} = 45, x_{11} = 10, x_{12} = 25, x_{13} = 35, x_{14} = 45, x_{15} = 40, x_{16} = 20, x_{17} = 30, x_{18} = 25, x_{19} = 20 \text{ \& } x_{20} = 10.$

$$\text{Mean} = \frac{\text{Sum of all the observations}}{\text{Total number of observations}}$$

$$\text{Therefore, } \bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + \dots + x_{18} + x_{19} + x_{20}}{20}$$

Where,

$$\sum_{i=1}^{20} x_i = \sum x = x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 + x_{10} + x_{11} + x_{12} + x_{13} + x_{14} + x_{15} + x_{16} + x_{17} + x_{18} + x_{19} + x_{20}$$

=

$$20 + 30 + 30 + 10 + 40 + 45 + 30 + 20 + 25 + 45 + 10 + 25 + 35 + 45 + 40 + 20 + 30 + 25 + 20 + 10$$

$$= 555$$

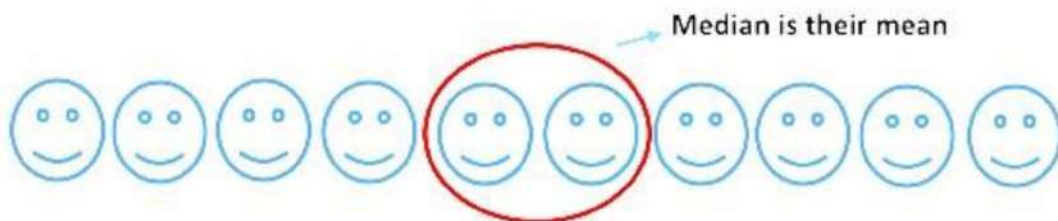
$$\text{Therefore, } \bar{x} = \frac{555}{20} = 27.75.$$

So, the mean of the marks obtained by 20 students of class 9th = 27.75.

Median: The median is the middle observation for a set of data that has been arranged in either ascending or descending order. Median is that observation that splits the arranged data into two halves.

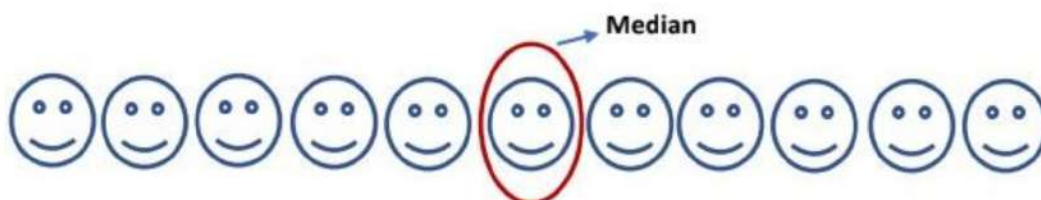
For an even and odd number of observations in ungrouped data, we have different approaches to find the median.

- When the number of observations (n) is odd, the median is the value of the $(\frac{n+1}{2})^{\text{th}}$ observation. For example, if $n = 11$, the value of the $(\frac{11}{2} + 1)^{\text{th}}$ is equal to 6th observation, which is the median as shown in the figure below.



- When the number of observations (n) is even, the median is the mean of the $(\frac{n}{2})^{\text{th}}$ and the $(\frac{n}{2} + 1)^{\text{th}}$ observations.

If $n = 10$ then the mean of the $(\frac{10}{2})^{\text{th}}$ and the $(\frac{10}{2} + 1)^{\text{th}}$ observations, i.e. the mean of the values of the 5th and 6th observations is the median, as shown in the figure below.



Example 2: The heights (in cm) of 11 students of a class 9th are as follows:

155, 160, 140, 130, 145, 135, 150, 152, 160, 142, 144.

Solution: First of all, we arrange the data in ascending order, as follows:

130, 135, 140, 142, 144, 145, 150, 152, 155, 160, 160.

Since the number of students is 11, an odd number, we find out the median by finding the height of the $(\frac{n+1}{2})^{\text{th}} = (\frac{11+1}{2})^{\text{th}} = \text{the } 6^{\text{th}} \text{ students}$, which is 145 cm. [where n is the number of students]


 Median

Example 3: The heights (in cm) of 10 students of a class 9th are as follows:

155, 160, 140, 130, 145, 135, 150, 152, 142, 144.

Solution: First of all, we arrange the data in ascending order, as follows:

130, 135, 140, 142, 144, 145, 150, 152, 155, 160.

Since the number of students is 10, an even number, the median is the mean of the $\left(\frac{n}{2}\right)^{\text{th}} = \left(\frac{10}{2}\right)^{\text{th}}$ and the $\left(\frac{n}{2} + 1\right)^{\text{th}} = \left(\frac{10}{2} + 1\right)^{\text{th}}$ observations. So, there are two middle terms, $\left(\frac{n}{2}\right)^{\text{th}} = 5^{\text{th}}$ and $\left(\frac{n}{2} + 1\right)^{\text{th}} = (5 + 1)^{\text{th}} = 6^{\text{th}}$. [where n is the number of students]

Therefore, the median is the mean of the values of the 5th and 6th terms.

$$\therefore \text{median} = \frac{144\text{cm} + 145\text{cm}}{2} = \frac{289}{2} \text{ cm} = 144.5 \text{ cm.}$$


 Median is their mean

Mode: It is defined as the most frequently occurring observations in data. That is, an observation with the maximum frequency is called the mode.



Example 4: The heights (in cm) of 12 students of a class 9th are as follows:

155, 160, 140, 130, 145, 135, 150, 152, 160, 142, 144, 160.

Solution: We can arrange the given data in ascending order:

130, 135, 140, 142, 144, 145, 150, 152, 155, 160, 160, 160.

Here 160 cm occurs most frequently, i.e. three times. So, the mode is 160 cm.

Example 5: The heights (in cm) of 12 students of a class 9th are as follows:

150, 160, 140, 144, 143, 153, 153, 155, 160, 160, 155, 155.

Solution: We can arrange the given data in ascending order:

140, 143, 144, 150, 153, 153, 155, 155, 155, 160, 160, 160.

Here, 155 cm and 160 cm all occur most frequently (three times).

So, the mode is 155 cm and 160 cm.

Summary of Statistics

Data

Data is a collection of facts, such as numbers, words, measurements, observations, etc. For example: The marks obtained by 10 students of class 9th in a mathematics test, out of 50 marks be:

39, 45, 33, 19, 21, 41, 21, 19, 40, 41. → This is data.

Types of data

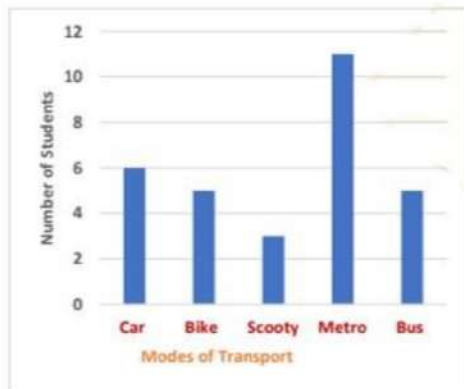
On the basis of the collection of fact On the basis of collection of data

- | | |
|------------------------------------|---|
| • Qualitative data | • Primary data |
| All the students passed the test. | Teacher prepared the result of the test. |
| • Quantitative data | • Secondary data |
| 10 students appeared for the test. | Principal asked for the result of the test. |

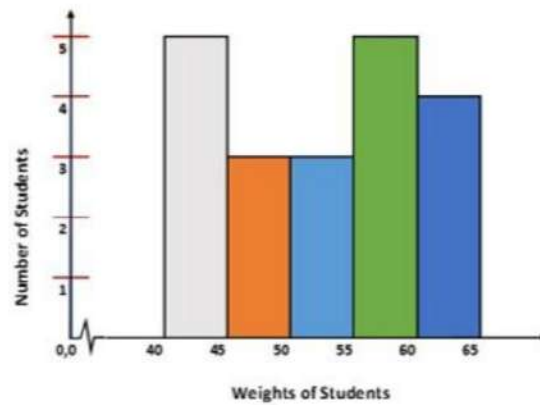
Quantitative data has also two types of data: Discrete & Continuous. In the example above, roll numbers from 1 to 10 follow a continuous data and collection of marks is called a discrete data.

Graphical representations of Data

Data can be represented graphically in the form of bar graphs, histograms, and frequency polygons.

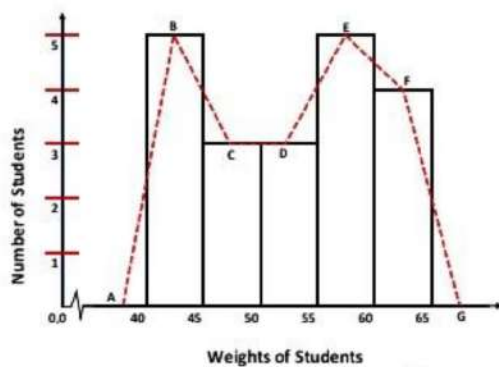


Bar Graph

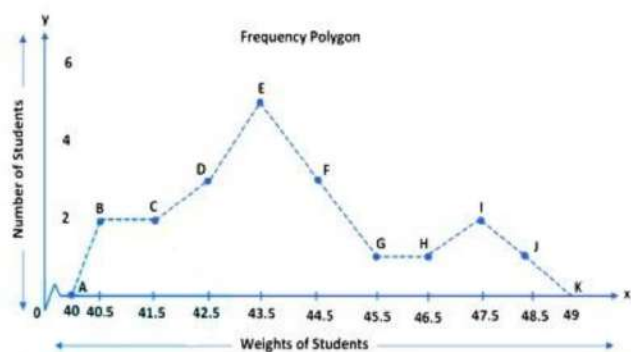


Histogram

Frequency polygons are shown in below



Frequency Polygon



Measure of Central tendency

Mean: It is calculated by dividing the sum of all values in a data by the number of values.

$$\text{Mean} = \frac{\text{Sum of all the observations}}{\text{Total number of observations}}$$

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Where, x_1, x_2, \dots, x_n , are observations and n is number of observations.

Mean of marks obtained by 10 students in mathematics test.

$$\begin{aligned}\text{Mean} &= \frac{39 + 45 + 33 + 19 + 21 + 41 + 21 + 19 + 40 + 41}{10} \\ &= \frac{319}{10} = 31.9\end{aligned}$$

Mode: It is defined as the most frequently occurring observations in a data. In above data, 19, 21 and 41 are the mode.

Median: It is the middle observation for a set of data that has been arranged in either ascending or descending order.

When the number of observations (n) is odd, the median is the value of the $(\frac{n+1}{2})^{\text{th}}$ observation.

When the number of observations (n) is even, the median is the mean of the $(\frac{n}{2})^{\text{th}}$ and the $(\frac{n}{2} + 1)^{\text{th}}$ observations.

19, 19, 21, 21, 33, 39, 40, 41, 41, 45.

Number of observations is 10, which is even.

$$\begin{aligned}\therefore \text{Median} &= \frac{(\frac{10}{2})^{\text{th}} + (\frac{10}{2} + 1)^{\text{th}}}{2} \\ &= \frac{(5)^{\text{th}} + (6)^{\text{th}}}{2} = \frac{33 + 39}{2} \\ &= 7.2\end{aligned}$$

Data

Data is a collection of facts, such as numbers, words, measurements, observations, etc. For example: The marks obtained by 10 students of class 9th in a mathematics test, out of 50 marks be:

39, 45, 33, 19, 21, 41, 21, 19, 40, 41. → This is data.

Types of data

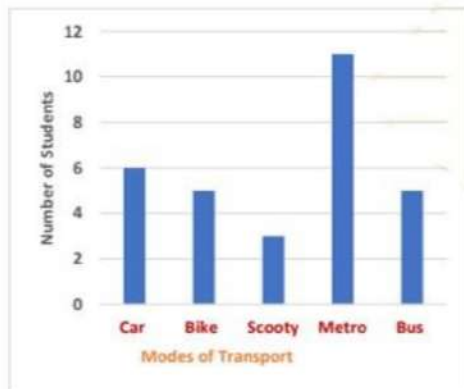
On the basis of the collection of fact On the basis of collection of data

- | | |
|------------------------------------|---|
| • Qualitative data | • Primary data |
| All the students passed the test. | Teacher prepared the result of the test. |
| • Quantitative data | • Secondary data |
| 10 students appeared for the test. | Principal asked for the result of the test. |

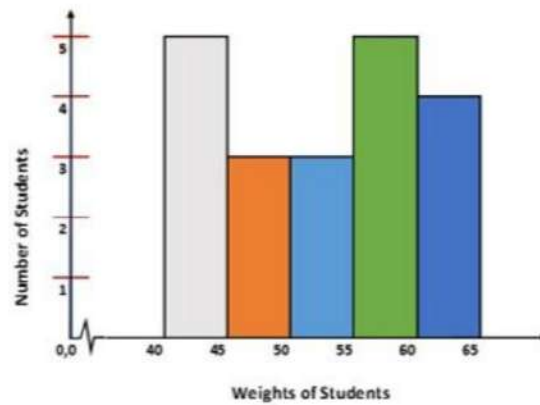
Quantitative data has also two types of data: Discrete & Continuous. In the example above, roll numbers from 1 to 10 follow a continuous data and collection of marks is called a discrete data.

Graphical representations of Data

Data can be represented graphically in the form of bar graphs, histograms, and frequency polygons.

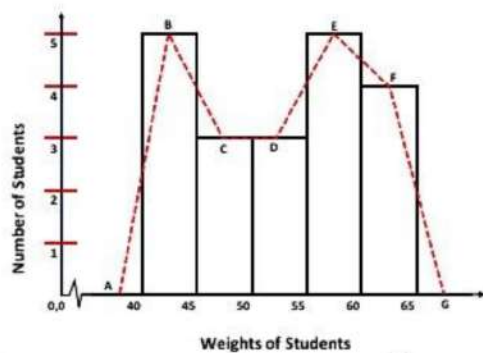


Bar Graph

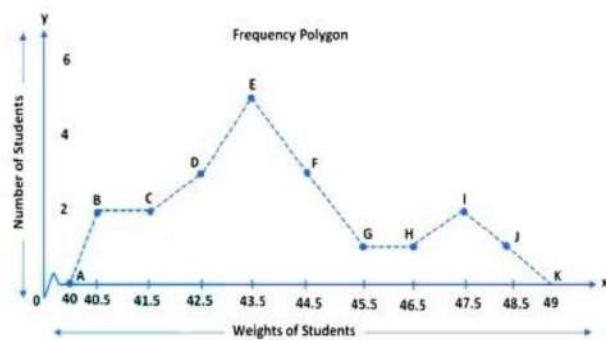


Histogram

Frequency polygons are shown in below



Frequency Polygon



Measure of Central tendency

Mean: It is calculated by dividing the sum of all values in a data by the number of values.

$$\text{Mean} = \frac{\text{Sum of all the observations}}{\text{Total number of observations}}$$

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Where, x_1, x_2, \dots, x_n , are observations and n is number of observations.

Mean of marks obtained by 10 students in mathematics test.

$$\begin{aligned}\text{Mean} &= \frac{39 + 45 + 33 + 19 + 21 + 41 + 21 + 19 + 40 + 41}{10} \\ &= \frac{319}{10} = 31.9\end{aligned}$$

Mode: It is defined as the most frequently occurring observations in a data. In above data, 19, 21 and 41 are the mode.

Median: It is the middle observation for a set of data that has been arranged in either ascending or descending order.

When the number of observations (n) is odd, the median is the value of the $(\frac{n+1}{2})^{\text{th}}$ observation.

When the number of observations (n) is even, the median is the mean of the $(\frac{n}{2})^{\text{th}}$ and the $(\frac{n}{2} + 1)^{\text{th}}$ observations.

19, 19, 21, 21, 33, 39, 40, 41, 41, 45.

Number of observations is 10, which is even.

$$\begin{aligned}\therefore \text{Median} &= \frac{(\frac{10}{2})^{\text{th}} + (\frac{10}{2} + 1)^{\text{th}}}{2} \\ &= \frac{(5)^{\text{th}} + (6)^{\text{th}}}{2} = \frac{33 + 39}{10} \\ &= 7.2\end{aligned}$$