

# STATISTICAL METHODS-1



Ms. Som Prakar

Department of Economics GC Kullu

	Title	Credits	
		L	T
I.	<b>INTRODUCTION TO STATISTICS</b> Statistics: Meaning, Scope, Nature, Function, Importance and Limitations of statistics. Types of Data: Primary and Secondary data, Univariate and Bivariate data, qualitative and quantitative data; nominal and ordinal data, Cross-section and Time Series. Sources of data: Primary and Secondary data. Diagrammatic and Graphic Presentation of Data	12	4
II.	<b>CENSUS AND SAMPLE</b> Collection of Statistical Data: Census and sample Method, Merits and demerits of census and sampling. Some basic sampling methods: Probability and Non Probability Sampling Methods with merits and demerits. Essentials of sampling, Methods of Selecting Sample, Sampling and Non- Sampling Errors.	11	4
III.	<b>MEASURES OF CENTRAL TENDENCY</b> Objectives of Averaging, Requisites of a Good Average. Arithmetic Mean, Median, Mode, Geometric Mean and Harmonic Mean. Quartiles, Deciles, Percentiles and Limitations of Averages.	10	4
IV.	<b>DISPERSION</b> <b>Meaning and significance of dispersion. Measures of dispersion: Range, Quartile Deviation, Mean Deviation, Standard Deviation, Coefficient of Variation, Variance, Absolute and Relative measures of variation - Lorenz Curve.</b>	11	4

## Origin of Statistics(सांख्यिकी का उदगम )

The word statistics is derived from the Latin word "status" or the Italian word "Statista," and meaning Of these words is "political state" or "government."

शासन को व्यवस्थित रूप से चलाने के लिए राज्य आँकड़े एकत्र करते थे जैसे सेवा की संख्या, रसद की मात्रा, कर्मचारियों का वेतन आदि। आंकड़ों की सहायता से राज्य के आय-व्यय का सही अनुमान लगाया जाता था। राजाओं की नीति बहुत अंशों तक आंकड़ों पर निर्भर करती थी इसलिए सांख्यिकी मूल रूप में राज्य के लिए उपयोगी विभिन्न पक्षों पर संख्यात्मक आंकड़ों का संग्रह

The original idea of "statistics" was the collection of information about and for the "state"

**Sir Ronald Fisher** (1890-1962) British statistician and biologist. is known as the father of modern statistic.

## **Prasanta Chandra Mahalanobis: Father of Indian statistics**

The word statistics has two different meanings.

**Statistics शब्द का दो अर्थों में प्रयोग होता है**

(1) Plural Sense(बहुवचन में )

(2) Singular Sense(एकवचन में)

**(1) Plural Sense: समक या आंकड़े, (Data),**

बहुवचन में संख्याकी का तात्पर्य समंको अथवा आंकड़ो से होता है जो किसी क्षेत्र में से सम्बन्धित संख्यात्मक मान होते हैं राष्ट्रीय आय के समंक(data), जनसंख्या के समंक(Data), उत्पादन के समंक(data) आदि।

In the plural sense, the word statistics refers to Information in terms of numbers and figures collected in a systematic manner with a definite purpose in any field of study.

**Singular Sense(एक वचन में) :** किन्तु जब इसका प्रयोग एकवचन में होता है तो सांख्यिकी विज्ञान से है जिसमें समंकों का संकलन, प्रस्तुतीकरण, विश्लेषण तथा निर्वचन से संबंधित क्रियाओं, अर्थात् सांख्यिकीय विधियों का अध्ययन किया जाता है।

In a singular sense, it refers to tools techniques or methods which are used in the collection, analysis, interpretation and presentation of numerical data.

इस प्रकार यह एकवचन और बहुवचन दोनों रूपों में प्रयोग होता है।

समंक(data)तथ्यों के समूह को व्यक्त करने वाले अंक होते हैं तथा सांख्यिकीय क्रियाएं इन्हीं डाटा (Data) को लेकर की जाती हैं जबकि इन समंकों (data)का संकलन, उनका अध्ययन तथा उनको सरल रूप में प्रस्तुत करना सांख्यिकी है।

# What is Statistics?

Statistics is a mathematical science of collecting, organizing, analyzing and interpreting data in such a way that meaningful conclusions can be derived from the data

सांख्यिकी, [गणित] की वह शाखा है जिसमें आँकड़ों का संग्रहण, प्रदर्शन, वर्गीकरण और उसके गुणों का आकलन का अध्ययन किया जाता है। सांख्यिकी एक गणितीय विज्ञान है जिसमें किसी वस्तु/अवयव/तंत्र/समुदाय से सम्बन्धित आँकड़ों का संग्रह, विश्लेषण, व्याख्या या स्पष्टीकरण और प्रस्तुति की जाती है।

Collection of tools and technique that are used to convert data into meaningful information

Statistics is science of dealing with numbers, facts and figures.

Statistics are based on data

## **आंकड़ों का संग्रह (Collection of Data)-**

प्रत्येक सांख्यिकीय अनुसंधान के लिए यह सबसे पहला व आवश्यक कार्य है। समस्या के अनुसार ही यह निश्चित किया जाता है कि कब, कहाँ से, किस ढंग से और कितने

## **(2) वर्गीकरण (Classification)-**

इकट्ठे किये हुए आंकड़ों को अधिक सरल व तुलना योग्य बनाने के लिए किसी गुण विशेष के आधार पर विभिन्न वर्गों में बाँटते हैं। वर्गीकरण विशेषतः वजन, रंग, स्थान आदि किसी भी गुण पर आधारित हो सकता है। यह संक्षिप्तीकरण की दिशा में एक कदम है।

## **(3) सारणीयन (Tabulation)-**

आंखों को अधिक अच्छा लगाने, मस्तिष्क में आसानी से बैठ जाने तथा तुलना योग्य बनाने के लिए वर्गीकृत आंकड़ों को और अधिक सरल व स्पष्ट बनाया जाता है और सारणी बनाकर उसमें शीर्षक लिखकर संख्याओं को विभिन्न खानों में भरा जाता है।



#### **(4) प्रस्तुतीकरण (Presentation)-**

इस रीति के द्वारा इकट्ठा किए हुए आंकड़ों को केवल आंखों को अच्छा लगने वाला ही नहीं बनाया जाता बल्कि ऐसा भी प्रयत्न किया जाता है कि बिन्दु रेखाओं या चित्रों द्वारा उन्हें प्रदर्शित किया जाय ताकि उनकी एक अमिट छाप मस्तिष्क पर पड़ जाय ।

#### **(5) विश्लेषण (Analysis)**

विभिन्न विधियों के द्वारा आंकड़ों का विश्लेषण किया जाता है और उनकी विशेषताएं ज्ञात की जाती हैं। ये विधियाँ हैं – माध्य या औसत, अपकिरण, विषमता तथा सहसम्बन्ध का मापन । इन विधियों के द्वारा आंकड़ों की परस्पर तुलना भी की जाती है।

#### **(6) निर्वचन (Interpretation)–**

विश्लेषणात्मक अध्ययन के पश्चात् निर्वचन की विधि के द्वारा परिणामों से निष्कर्ष निकालते हैं।



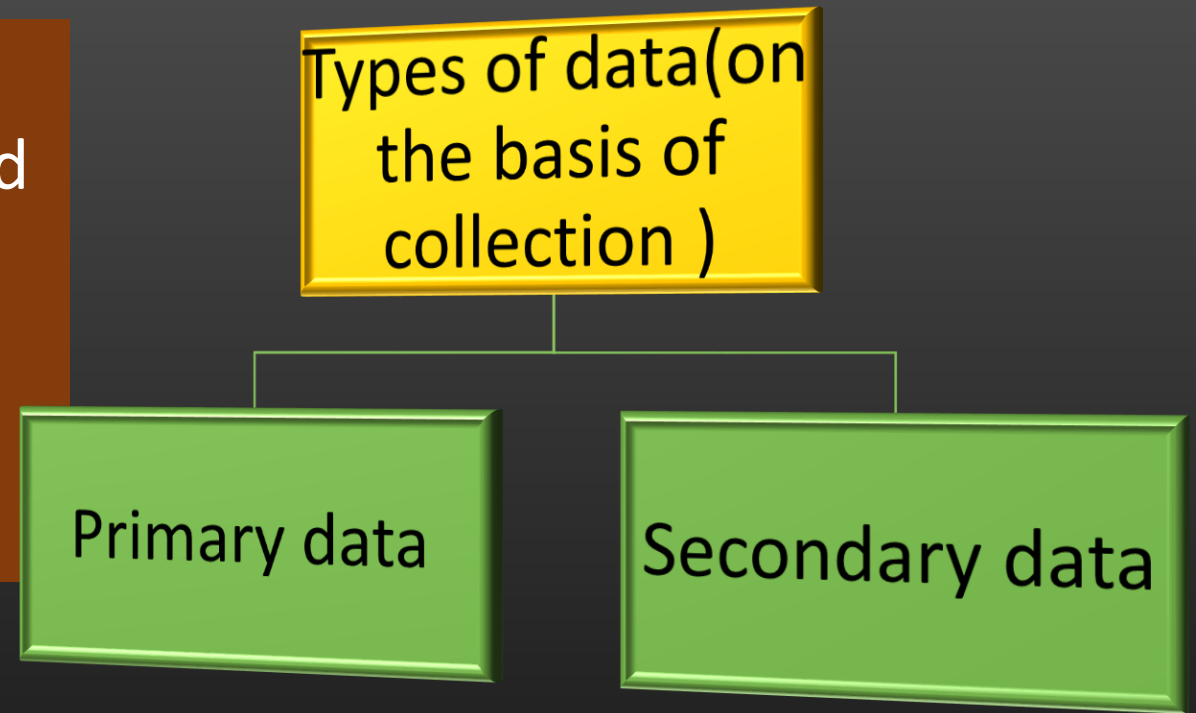
Data( आँकड़े ):

A collection of facts and information (raw information facts)

- It is a collection of facts and figures to be used for a specific purpose such as a survey or analysis.
- E.g. no of students in a class, no of schools in a city ,heights of the students in your class etc.
- When arranged in an organized form, can be called information.

Primary data(प्राथमिक आँकड़े ) :means original /pure data that has been collected for the first time from original source for specific purpose An example of primary data is the Census of India.

Secondary data(द्वितीय आँकड़े ) : are those which are already in existence which have been collected for some other purpose .collected by others .Data that is being reused e.g. data from books newspapers magazines internet etc.



OtherTypes of  
data

Qualitative data  
(गुणात्मक आँकड़े)

Quantitative  
data(मात्रात्मक  
आँकड़े )

Nominal  
data

Ordinal data

Discrete

Continuous

Qualitative data  
/categorical  
data

Nominal

Ordinal

Qualitative Data { गुणात्मक आँकड़े } :  
जिन्हें संख्याओं में नहीं दर्शा सकते

Data is qualitative when data measures a quality and characteristics

They may be observed but cannot be computed or calculated.

For example, data on attributes such as intelligence, honesty, wisdom, cleanliness, and creativity .

Nominal Data (नामवाचक आँकड़े ):  
information organized by category or name

Gender : Male Female

Colour Of Hair: Black .Brown ,Blonde

Ordinal data (क्रमवाचक आँकड़े );  
information organized in particular order  
ranking /meaningful orders

1<sup>st</sup> rank 2<sup>nd</sup> rank 3<sup>rd</sup> rank

delighted Satisfied Non satisfied

Quantitative Data/numerical data(मात्रात्मक आँकड़े ) :  
जिन्हे संख्याओं में दर्शाया जा सके )These can be measured .

They can be numerically represented and calculated .

For example,

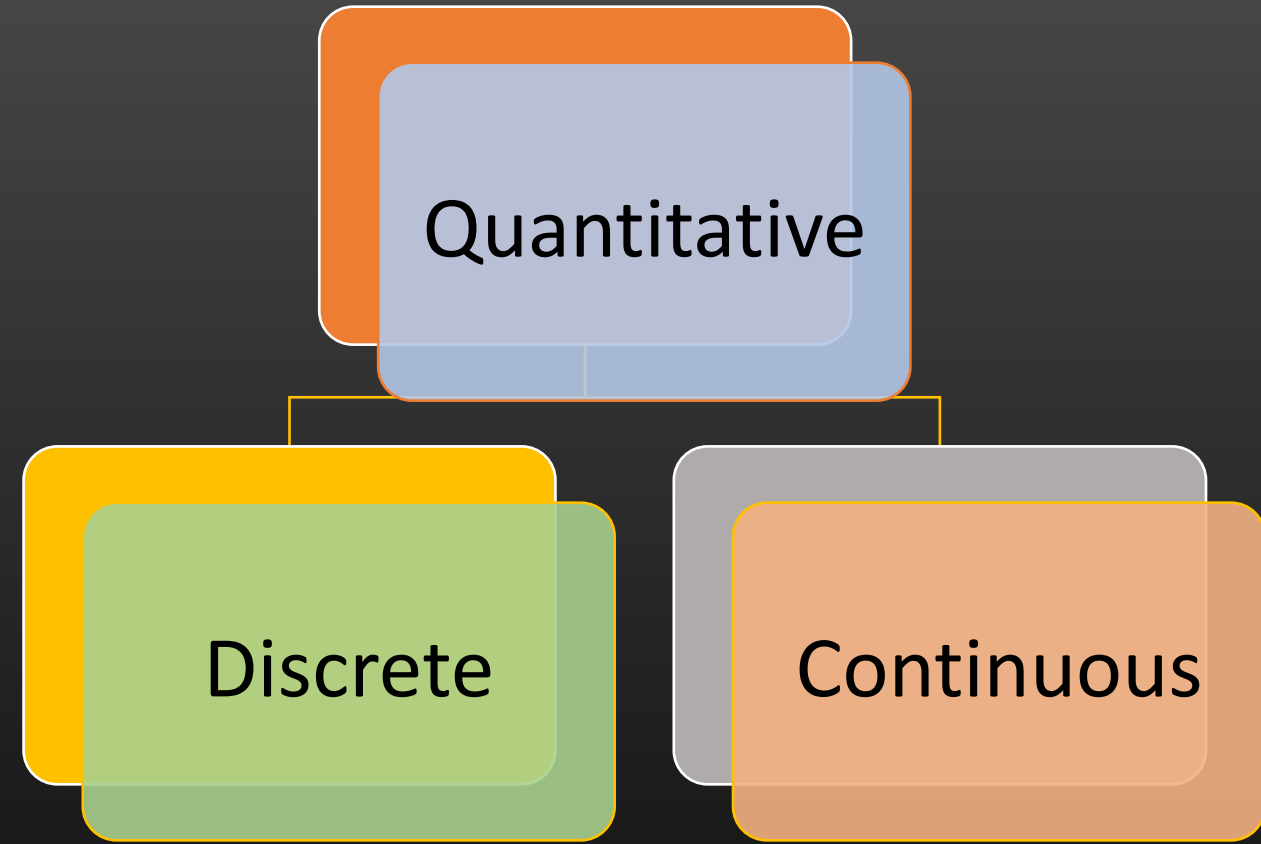
Height:5 feet

temp:30 degrees

Weight:55 kg

number of children:2

Age= 25yrs



Discrete  
data(खण्डित  
आँकड़े ) &  
Continuous data  
(अखण्डित आँकड़े )

Discrete : that can  
be counted  
( in whole no s )  
No of students in  
class :4, 20  
No of cars :  
No of trees :

- Continuous : can take any numerical value within a range
- Height Weight
- Shoe size



## Univariate Data(एक चर आँकड़े )

Univariate means "one variable" (one type of data) As the name suggests, "Uni," meaning "one," in univariate analysis, there is only one dependable variable.

## Bivariate data(द्विचर आँकड़े ) –

This type of data involves two different variables. The analysis of this type of data deals with causes and relationships and the analysis is done to find out the relationship among the two variables



# Cross-sectional data( :

refers to a set off observations taken at a single point in time

Data on different entities for a single time period

Different entities: Entities could be individuals, households, firms, states, countries and so forth. In case of cross sectional data, entities are allowed to vary.

There is single time period for which data has been collected.

For example, suppose we study the GDP of 3 developing countries for year 2019

Countries	GDP(2019 )
India	2trllion
Nepal	1trillion
Srilanka	1.5trillion

Firms	Sales (August)
A	20
B	15
C	28
D	30

Time-series data refers to observations made over a period of time at regular intervals.

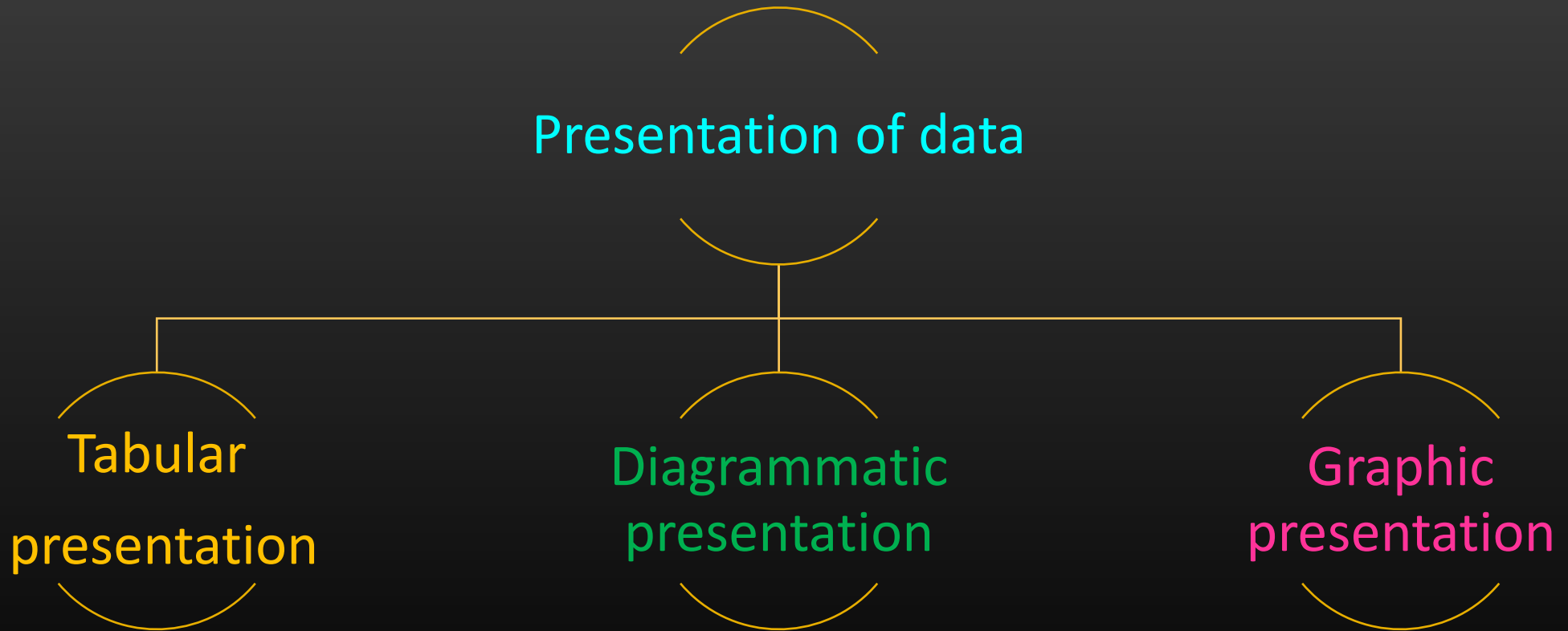
Time-series data contains observations on a single phenomenon over multiple periods of time.

e.g. sales of a firm in different months or different years

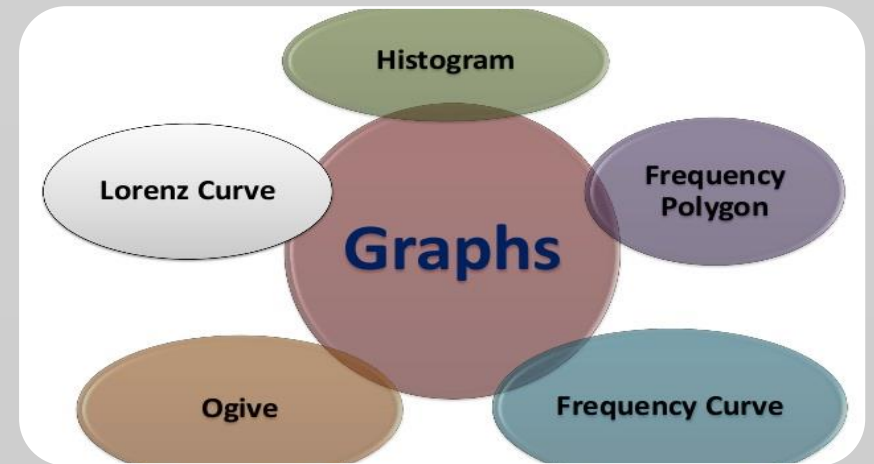
Months/years	Sales of firm A
2015	20
2016	28
2017	25
2018	20

# PRESENTATION OF DATA:

Method by which people organise ,summarise and communicate information using a variety of tools such as tables, graphs and diagrams



It is a systematic and logical arrangement of classified data in rows and columns.



## Diagrammatic Presentation

### Bar Diagram

Simple bar diagram -  
Multiple bar diagram

Pie diagram pictogram  
cartogram

### Histogram

Frequency Polygon

Cumulative

Frequency curve  
or Ogives

# Diagrammatic Presentation of Data

Diagrammatic presentation is a technique of presenting numeric data through Bar Diagrams ,Pie Diagrams, Pictograms & Cartograms etc.

## Bar Diagrams:

In which data are presented in form of bars or rectangles .

Bar Diagrams are rectangular in shape placed on the same base.

Their height represents the magnitude/value of the variable.

Width of all the bars and gap between the two bars is kept the same.

Bars can be vertical or horizontal

### Types of Bar Diagram

Simple bar diagram –

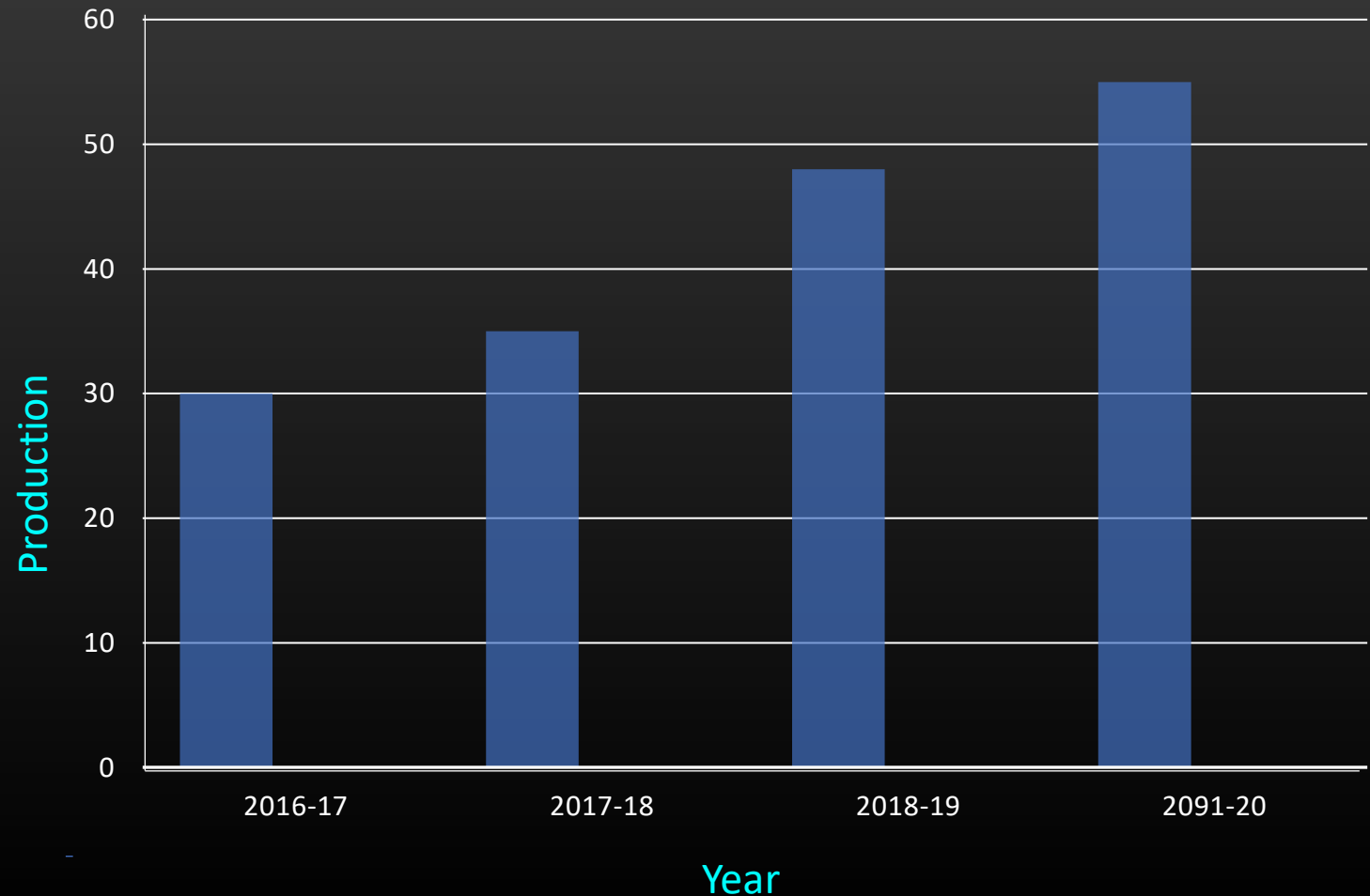
Multiple bar diagram

# Simple Bar Diagram:

Simple Bar diagram comprises of a group of rectangular bars of equal width for each class or category of data.. All bars stand on the same baseline. The bars are separated from each others by equal intervals.

Vertical bar

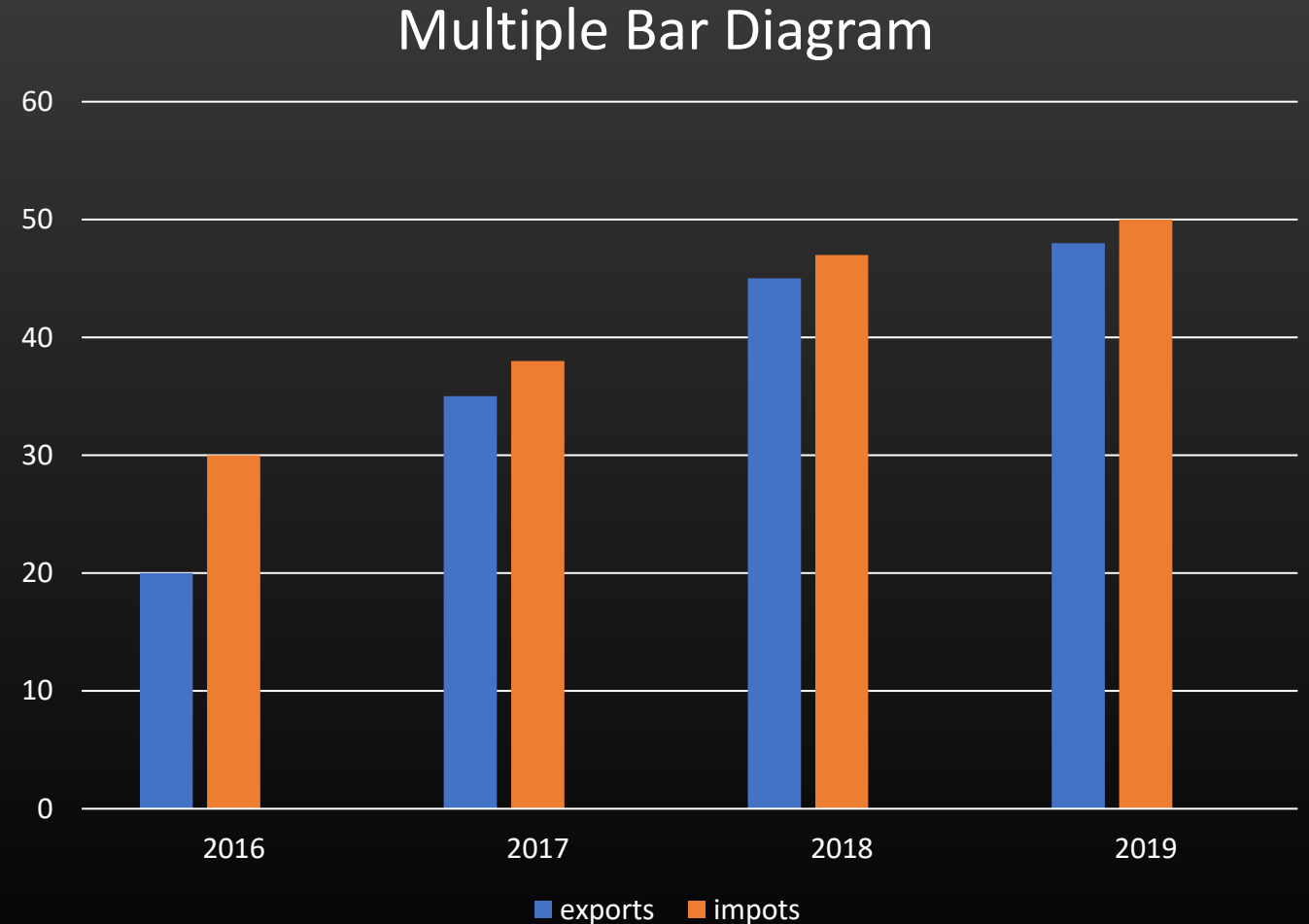
Year	Production (in tons )
2016-17	30
2017-18	35
2018-19	48
2019-20	55



# Multiple Bar Diagram

This diagram is used when we have to make a comparison between two or more variables like income and expenditure, import and export for different years, marks obtained in different subjects in different classes

Year	Exports (worth cr)	Imports
2016	20	30
2017	35	38
2018	45	47
2019	48	50





## Pie Chart or circular Diagram ?

A Pie Chart is a type of graph that displays data in a circular graph. The pieces of the graph are proportional to the fraction of the whole in each category. In other words, **each slice of the pie is relative to the size of that category** in the group as a whole. The entire “pie” represents 100 percent of a whole, while the pie “slices” represent portions of the whole.

## Pie or Circular Diagrams

In addition to bar diagrams, pie diagrams are also widely used to pictorially represent data. In this, a circle is divided into various segments which are decided on the basis of percentages. Which means the circle is divided into sectors depending on various percentages.

The steps for construction of a pie diagram are:

The first step involves finding out respective percentages. This is done by a simple mathematical formula to find out percentages which is –

*$\{(Parts\ for\ the\ respective\ sector)/total\ parts\} \times 100\}$ .*

For example, if in a class of 50 students, 5 are obese, 20 are fat and 25 are slim then the percentages will be as follows:

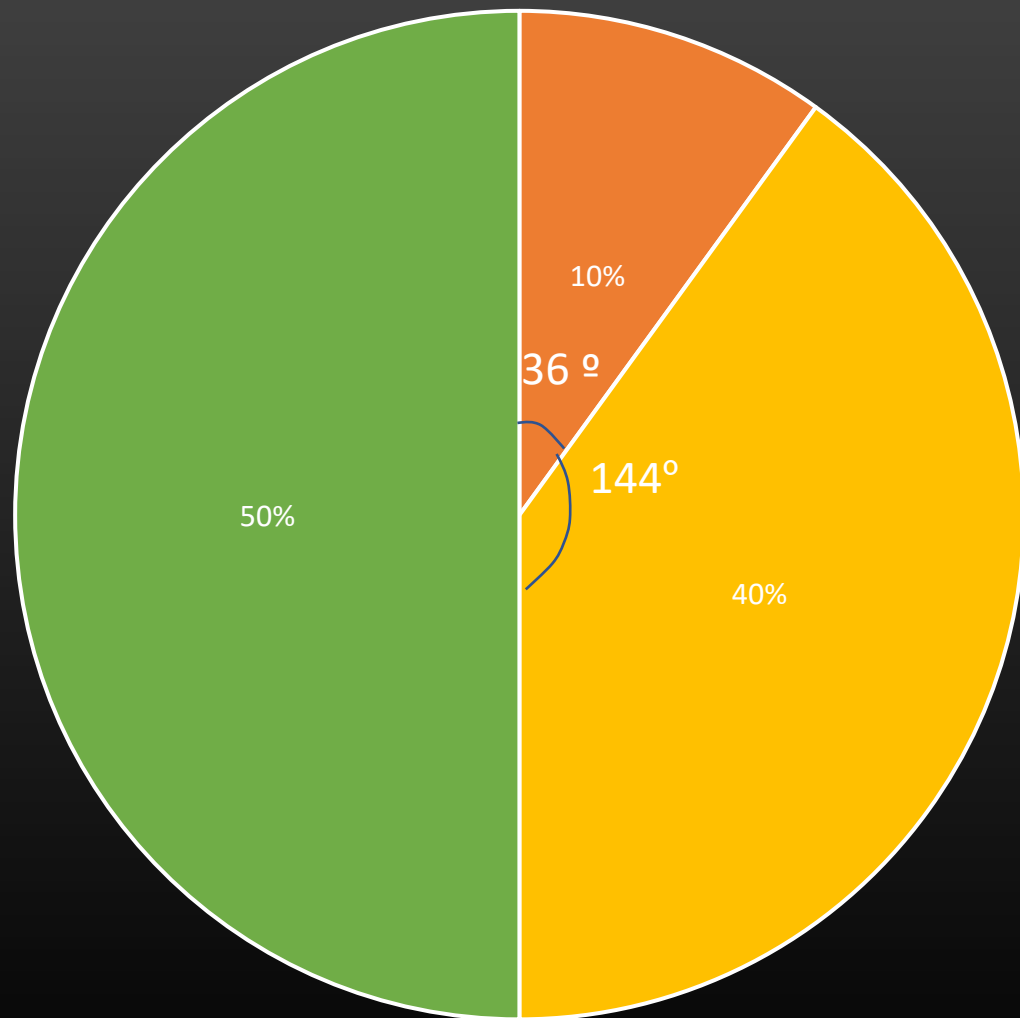
$$(5/50) \times 100 = 10\%$$
$$(20/50) \times 100 = 40\%$$
$$(25/50) \times 100 = 50\%$$

A circle comprises 360 degrees. The angles that each sector will span across is decided by the given formula

$$(Percentage\ value / 100) \times 360^\circ$$

100%	360°
1%	360/100=3.6°
10%	24/100 x360= 36 °
40%	144°
50 %	180°

Finally, just plot these values according to their respective angles on a circle and give appropriate markings to complete the pie chart.



■ obese ■ fat ■ slim ■

# Histogram:

**Histograms** are similar to bar charts;

It is graphic presentation of frequency distribution of continuous series

---

## Scores of 25 students

---

15,18,11,10,16

20,24,23,21,27,28,29,22,21.5, 22.5

36,35 ,34,38,36

43,48,49

68,67,6

---

Marks	Frequency	Scores Included
10-20	5	15,18,11,10,16
20-30	10	20,24,23,21,27,28,29,22,21.5, 22.5
30-40	5	36,35 ,34,38,36
40-50	3	43,48,49
60-70	2	68,67

# Histogram:

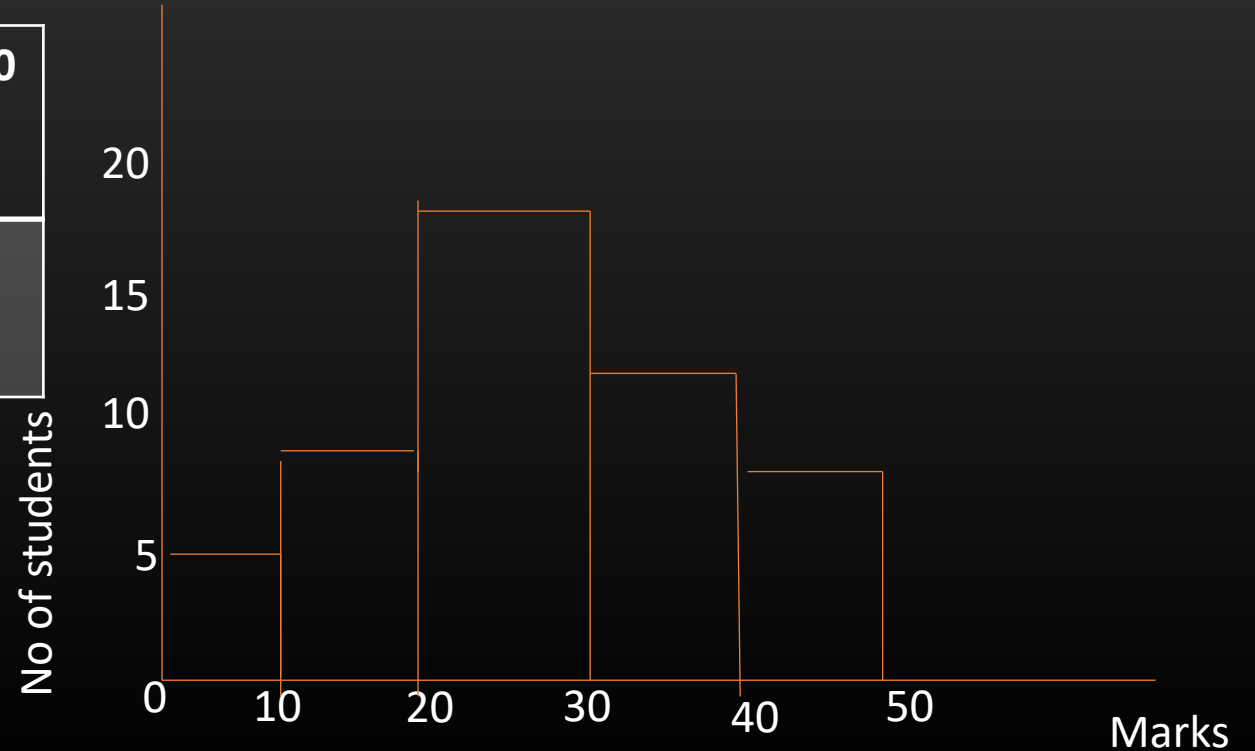
**Histograms** are similar to bar charts;

It is graphic presentation of frequency distribution of continuous series ,while constructing a histogram values of variable are shown on x-axis and frequency on y-axis

Histogram frequency distribution are of two types

1. Histogram of equal class intervals:

Marks	0-10	10-20	20-30	30-40	40-50
Students	5	10	20	12	8



## Histograms with Unequal Intervals

For histograms, the heights of the bars on a histogram indicate the frequency density. The area of each bar is the frequency. The Frequency Density changes for each bar on the histogram:

Frequency Density = frequency / class width

Remembering the formula for the area of a rectangle, it might be easier to remember:

Frequency (the area) = Class Width x Frequency Density

and rearrange to suit the requirements.



for example

Weekly wage (₹)	10 – 15	15 – 20	20 – 30	30 – 40	40 – 60
No of workers	6	4	12	18	8

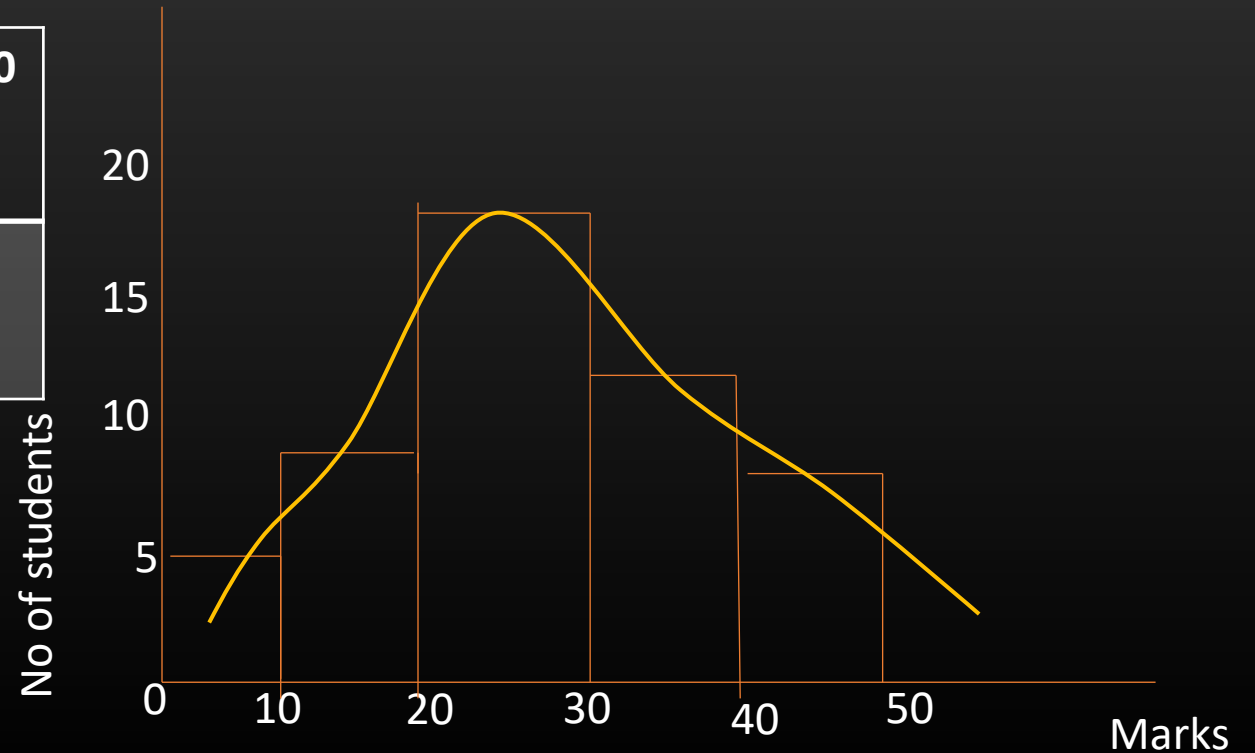
Adjustment factor/class width = class interval of concerned class

Lowest class interval

Weekly wage (₹)	10 – 15	15 – 20	20 – 30	30 – 40	40 – 60
No of workers	6	4	12	18	8
Class width	5	5	10	10	20
Adjustment factor	$5/5=1$	$5/5=1$	$10/5=2$	$10/5=2$	$20/5=4$
Frequency density	6	4	$12/2=6$	$18/2=9$	$8/4=2$

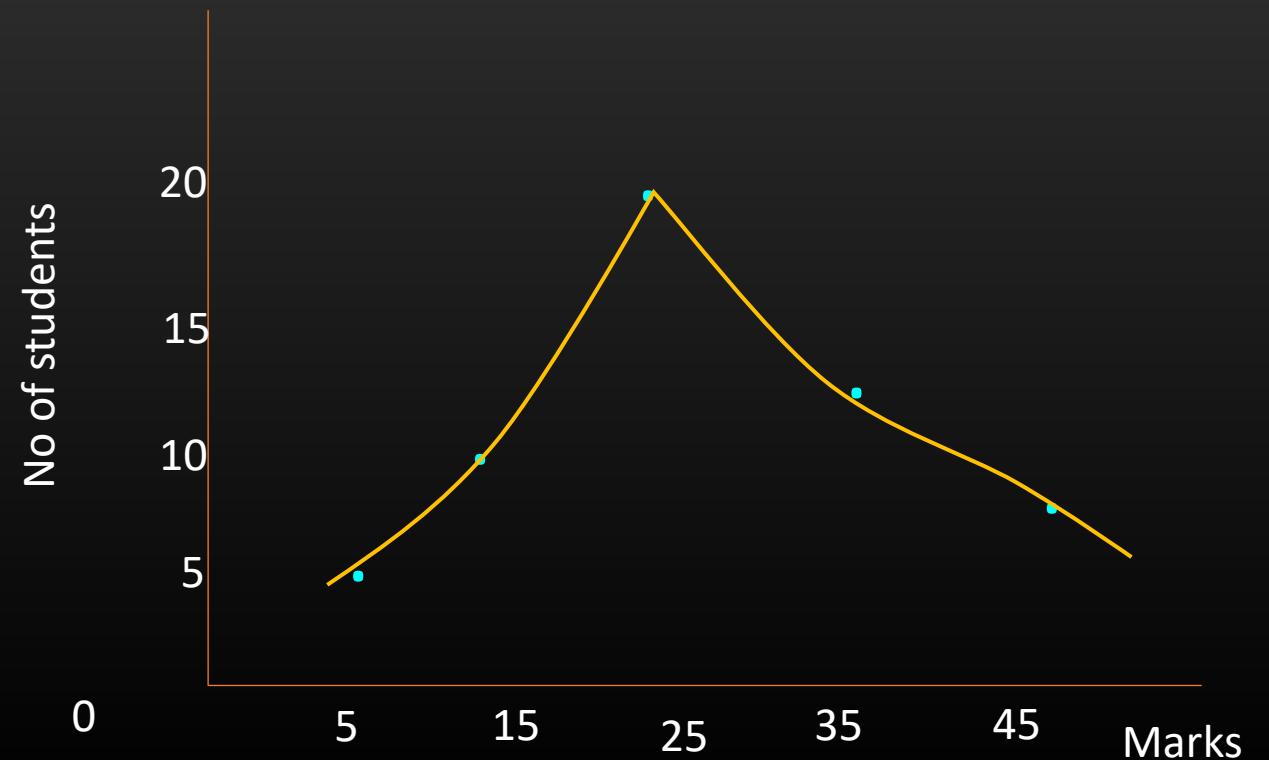
**Frequency Polygon** :A **frequency polygon** is a graph constructed by using lines to join the midpoints of each interval,. The heights of the points represent the frequencies. A frequency polygon can be created from the histogram or by calculating the midpoints of the bins from the frequency distribution table. The **midpoint** of a bin is calculated by adding the upper and lower boundary values of the bin and dividing the sum by 2.

Marks	0-10	10-20	20-30	30-40	40-50
Students	5	10	20	12	8



**Frequency Polygon** :A frequency polygon is a graph constructed by using lines to join the midpoints of each interval,. The heights of the points represent the frequencies. A frequency polygon can be created from the histogram or by calculating the midpoints of the bins from the frequency distribution table. The **midpoint** of a bin is calculated by adding the upper and lower boundary values of the bin and dividing the sum by 2.

Marks	0-10	10-20	20-30	30-40	40-50
Mid – point	5	15	25	35	45
Students	5	10	20	12	8



# Ogive/Cumulative Frequency Curve

An **Ogive** Chart is a **curve** of the cumulative frequency distribution or cumulative relative frequency distribution.

The graphs of the frequency distribution are frequency graphs that are used to exhibit the characteristics of discrete and continuous data. Such figures are more appealing to the eye than the tabulated data..

The two methods of Ogives are:

- Less than Ogive
- Greater than or more than Ogive

# ❑ Less than Ogive: convert simple frequency distribution to less than cumulative frequency distribution

## Simple frequency distribution

Marks	No of students (frequency )
0-10	5
10-20	10
20-30	20
30-40	10
40-50	5

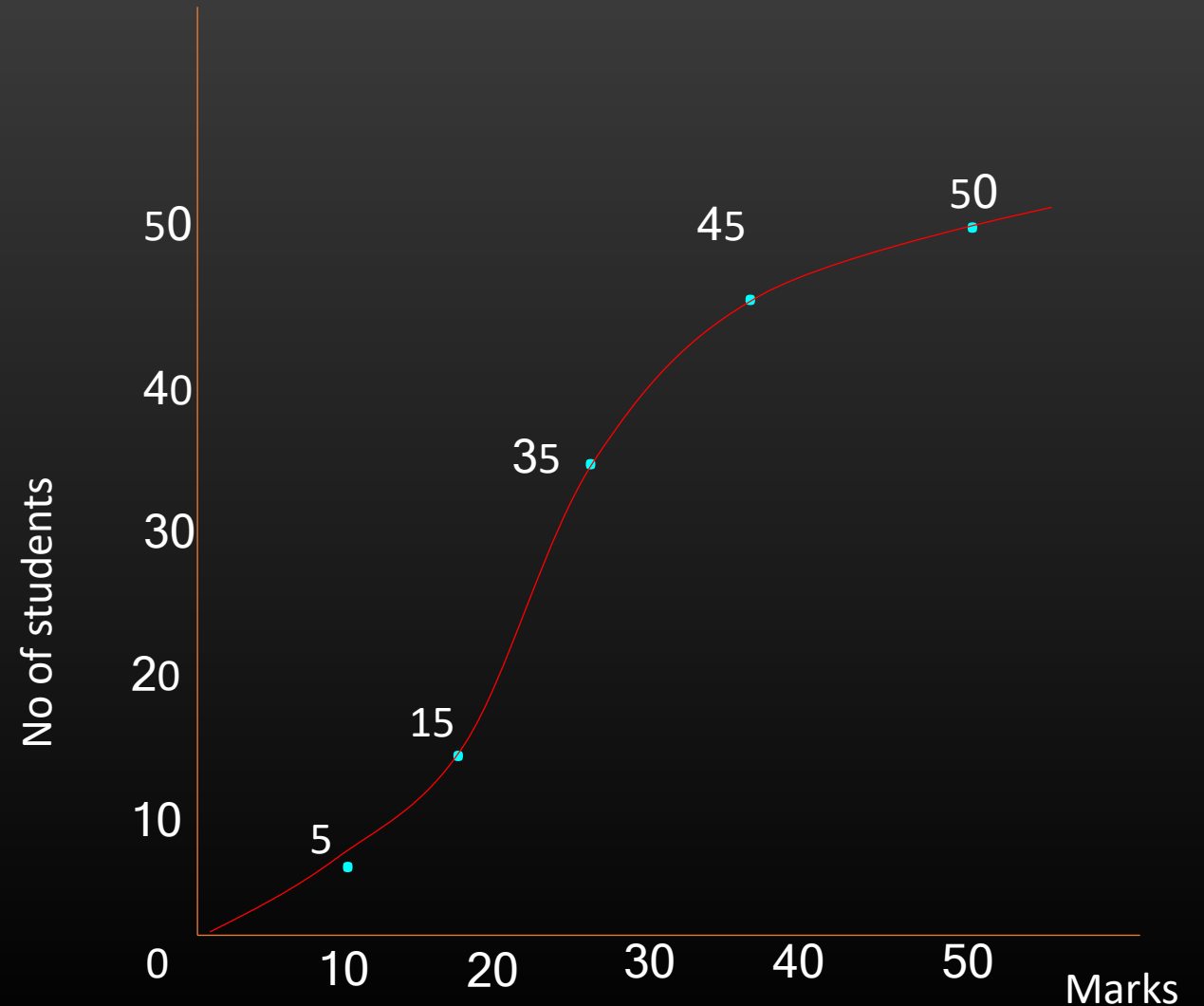
## Cumulative frequency distribution

Marks	No of students Cumulative frequency
Less than 10	5
Less than 20	15
Less than 30	35
Less than 40	45
Less than 50	50

## ❑ Less than Ogive:

Cumulative frequency distribution

Marks	No of students Cumulative frequency
Less than 10	5
Less than 20	15
Less than 30	35
Less than 40	45
Less than 50	50



# More than Ogive

## Simple frequency distribution

Marks	No of students (frequency )
0-10	5
10-20	10
20-30	20
30-40	10
40-50	5

## Cumulative frequency distribution

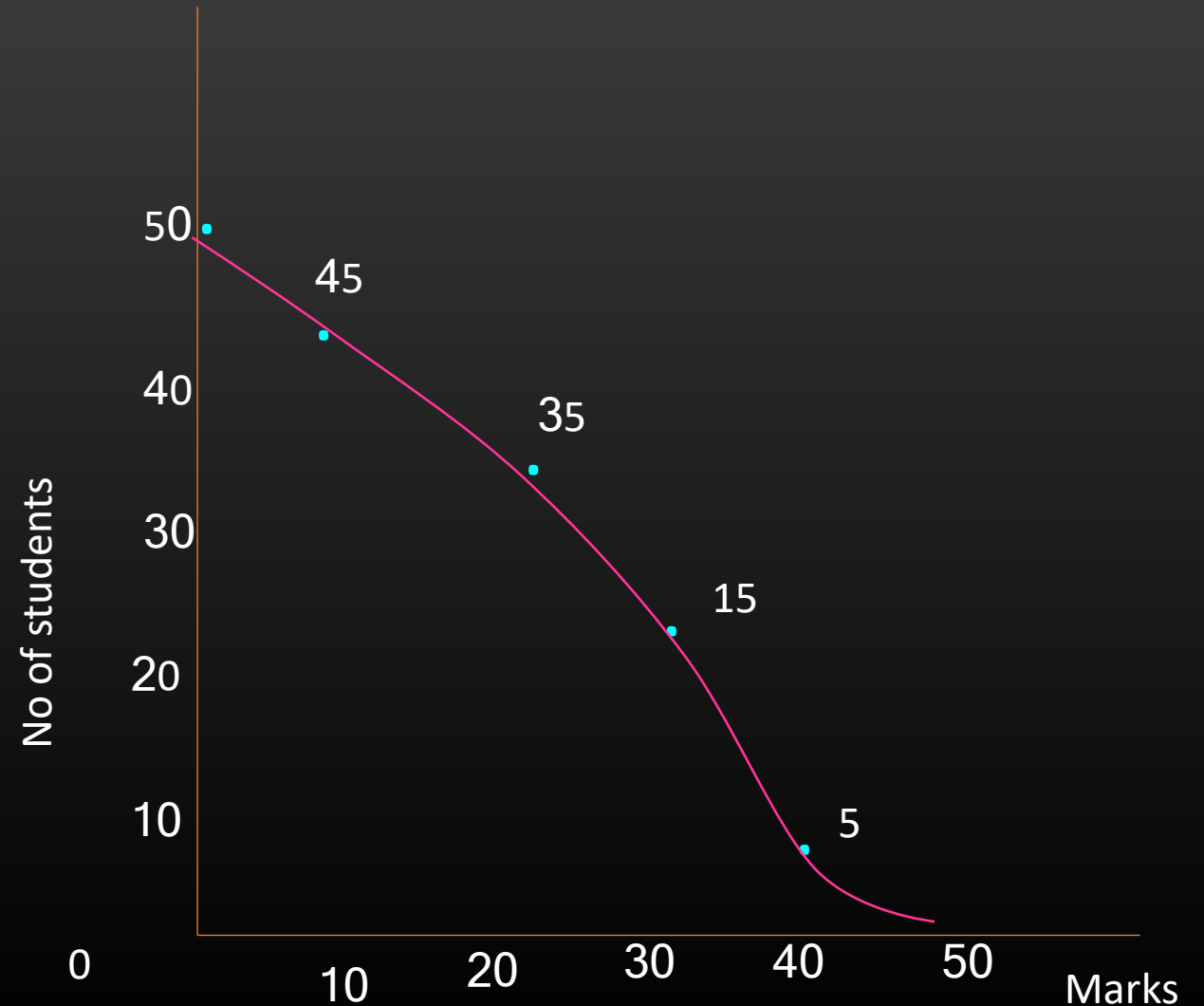
Marks	No of students Cumulative frequency
More than 0	50
More than 10	45
Less than 20	35
Less than 30	15
Less than 40	5



## ❑ Less than Ogive:

Cumulative frequency distribution

Marks	No of students Cumulative frequency
More than 0	50
More than 10	45
Less than 20	35
Less than 30	15
Less than 40	5



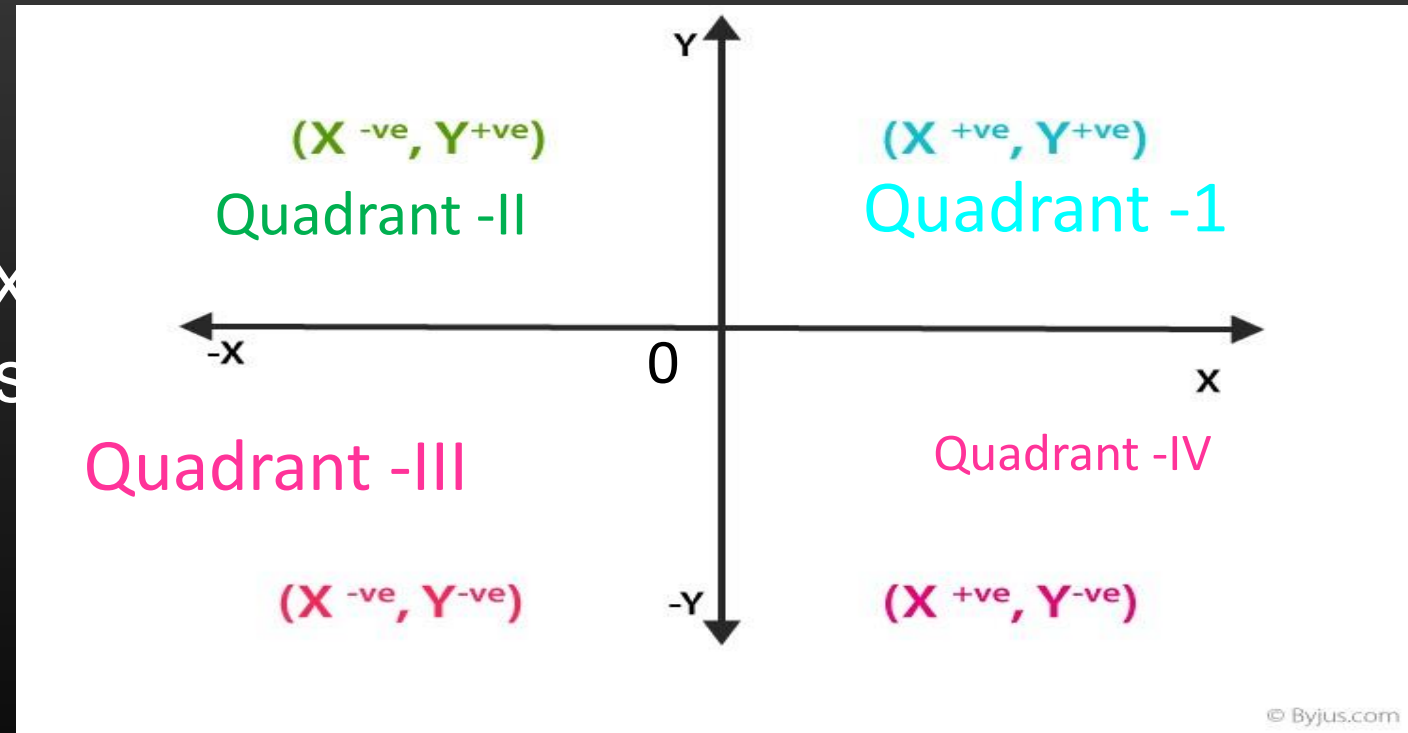
# Arithmetic line –graphs: Time Series Graphs

When data are presented on graph paper

An arithmetic line graph is also called time series graph. In this graph time/hour/day/date, week, month, year) is plotted along the X axis and the corresponding values of the variable along the Y-axis.

Arithmetic line graphs, along or time series graphs can

- (i) One variable graphs.
- (ii) Two or more than two variable



One variable graph Time series graph: Time is always shown on the horizontal axis.

year	Profit ('000₹)
2013	60
2014	72
2015	75
2016	65
2017	80
2018	95

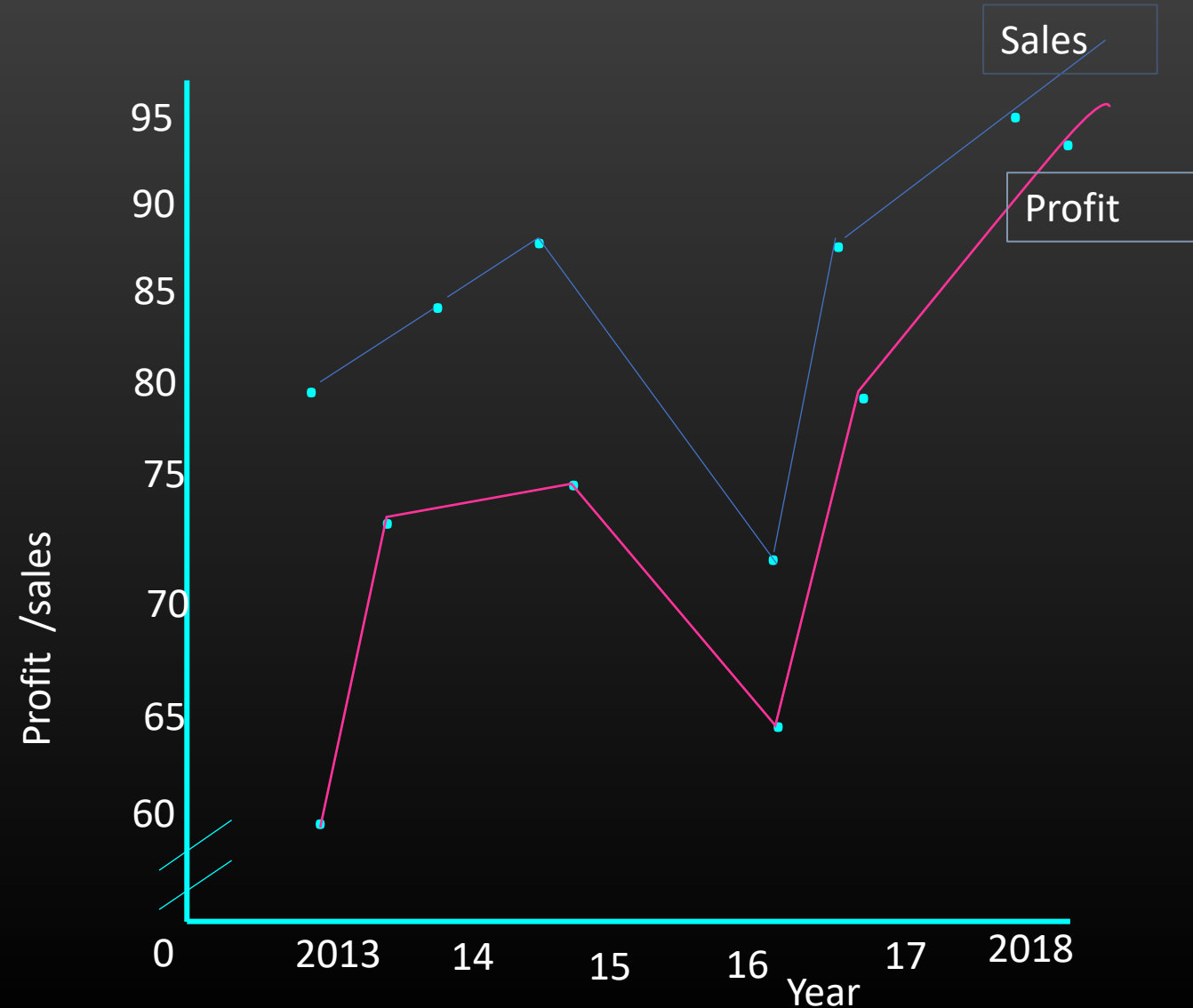
False base line is used when the values in a series are very large and the difference between the smallest value and zero is high.

False base line



Two or more than two variable graphs Time series graph: Time is always shown on the horizontal axis.

year	Profit ('000₹)	Sales (000₹)
2013	60	80
2014	72	85
2015	75	87
2016	65	70
2017	80	85
2018	95	90



# Data Collection in Statistics

In Statistics, the basis of all statistical calculations or interpretation lies in the collection of data. There are numerous methods of data collection. we shall focus on two primary methods and understand the difference between them. Both are suitable in different cases and the knowledge of these methods is important to understand when to apply which method.

These two methods are the **Census method and Sampling method**.

**Population or universe** : The population or universe represents the entire group of units which is the focus of the study

## Census Method

Census method when all members of the population are studied. A population refers to the set of all observations under concern. For example, if you want to carry out a survey to find out student's feedback about the canteen/library facilities of your college, all the students of your college would form a part of the 'population' for your study.

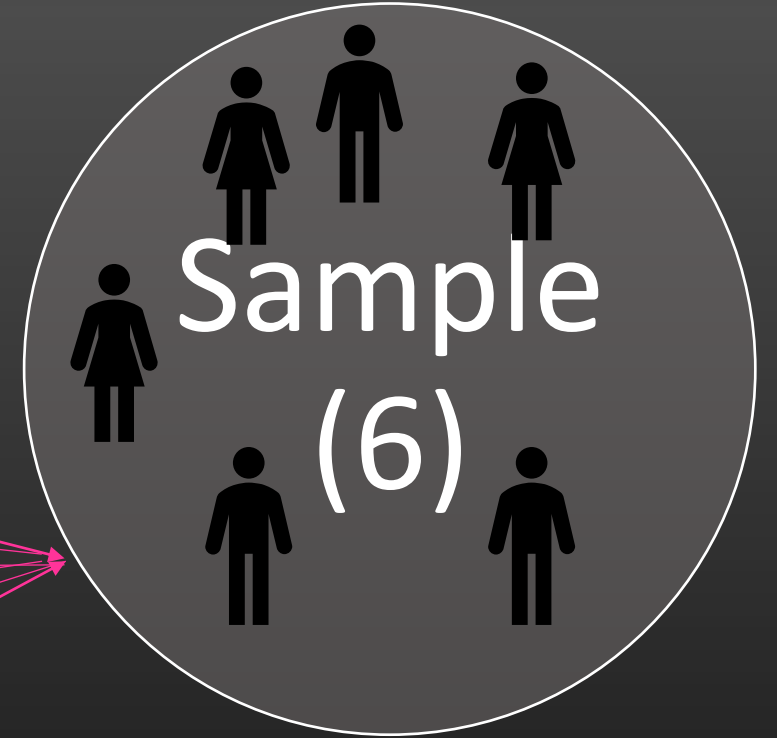
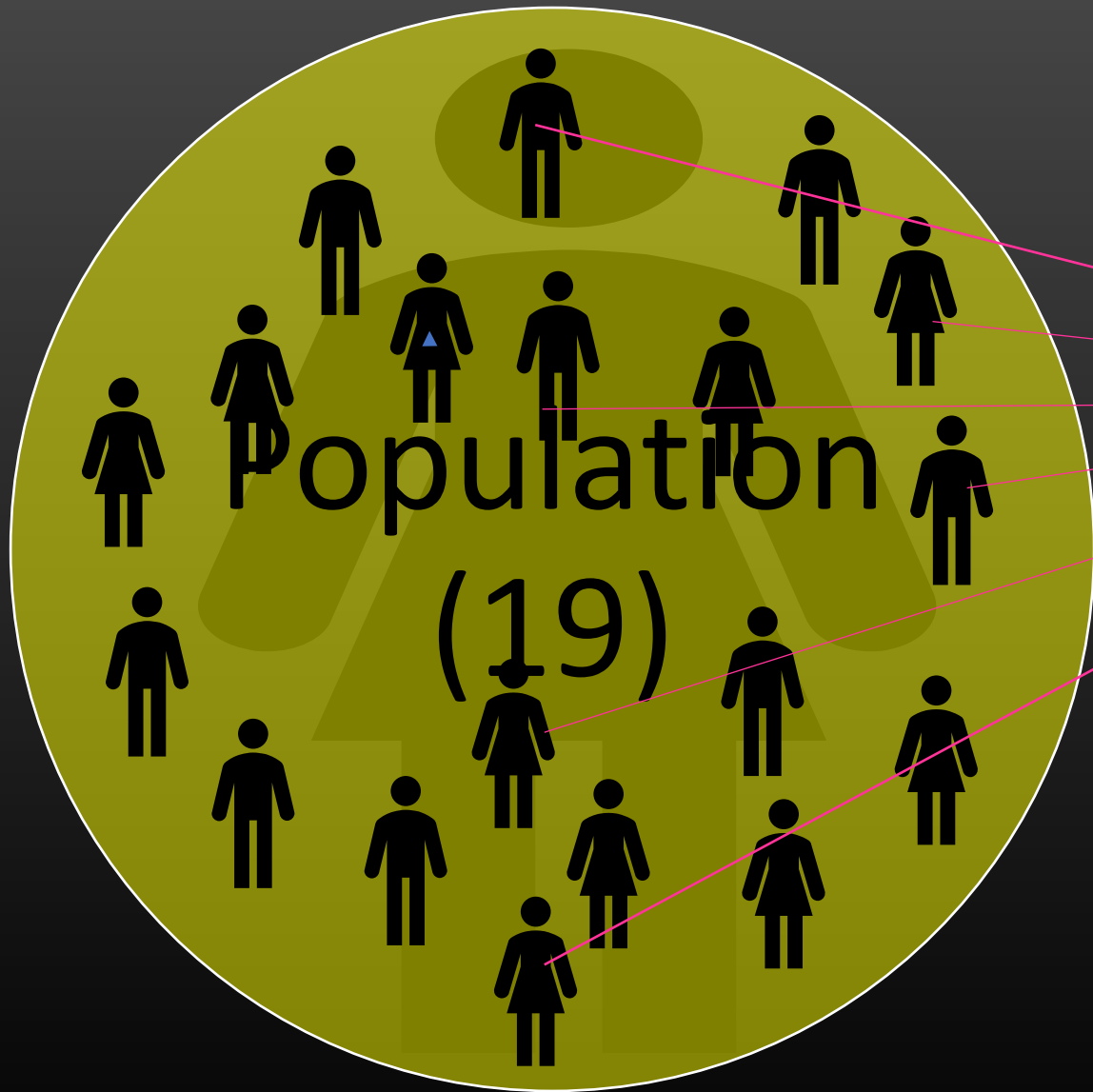
E.g . Population census Government conducts the **Census of India** every ten years. The Census appropriates information from households regarding their incomes, the earning members, the total number of children, members of the family, etc.

# Sampling Method

.But it is not always practical to collect information from all the units of the population.

It is a time-consuming and costly method. Thus, an easy way out would be to collect information from some representative group from the population and then make observations accordingly. This representative group which contains some units from the whole population is called the **sample**. In sample survey, information is collected from a few selected unit of the population.

For example, if you want to carry out a survey to find out student's feedback about the canteen/library facilities of your college , Instead of all the students of college we will select some students form a part of the 'population' for your study.





# Types of Sampling Methods - probability &

