

Statistical Description of Data

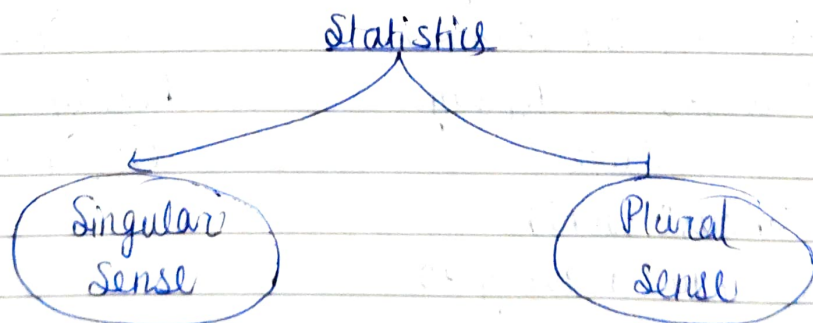
→ Statistics - A branch of mathematics dealing with the collection, analysis and interpretation/presentation of numerical data.

★ Statistical Survey

1. Collection of Data
2. Classification of Data
3. Representation of Data
4. Analysis of Data
5. Conclusion

Origin of Statistics word

- ★ latin word - Status (according to latin scholars)
- ★ German word - Statistik
- ★ Italian word - Statista
- ★ French word - ~~Statistique~~ Statistique



As a method / technique / science

★ Collection of data

★ Data

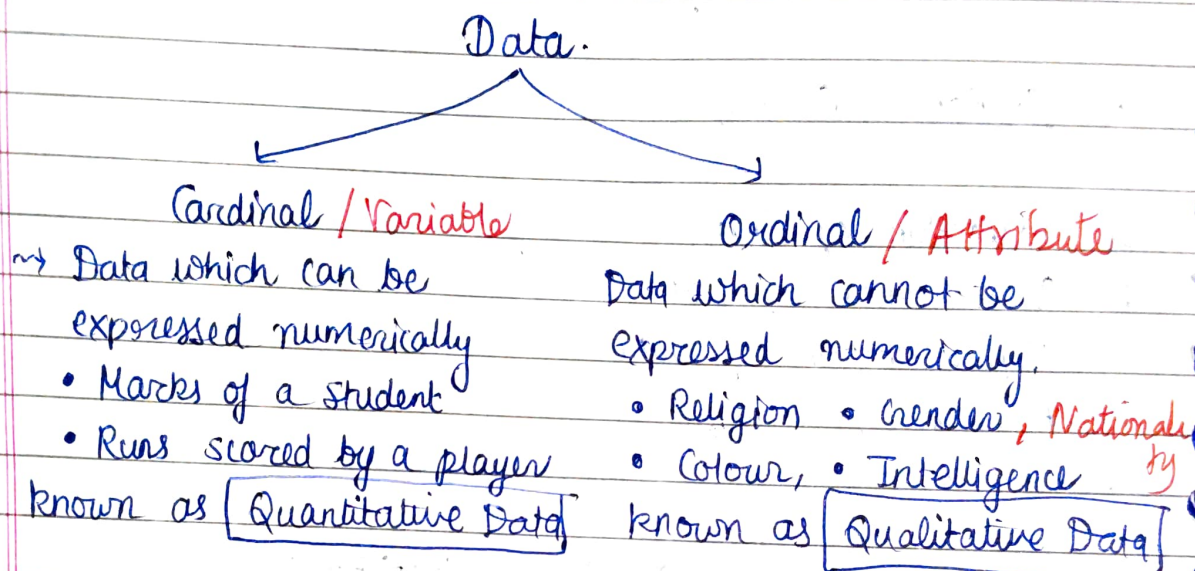
★ Representation of data

known as "Science of Data" - "Science of Averages" - "Science of counting"

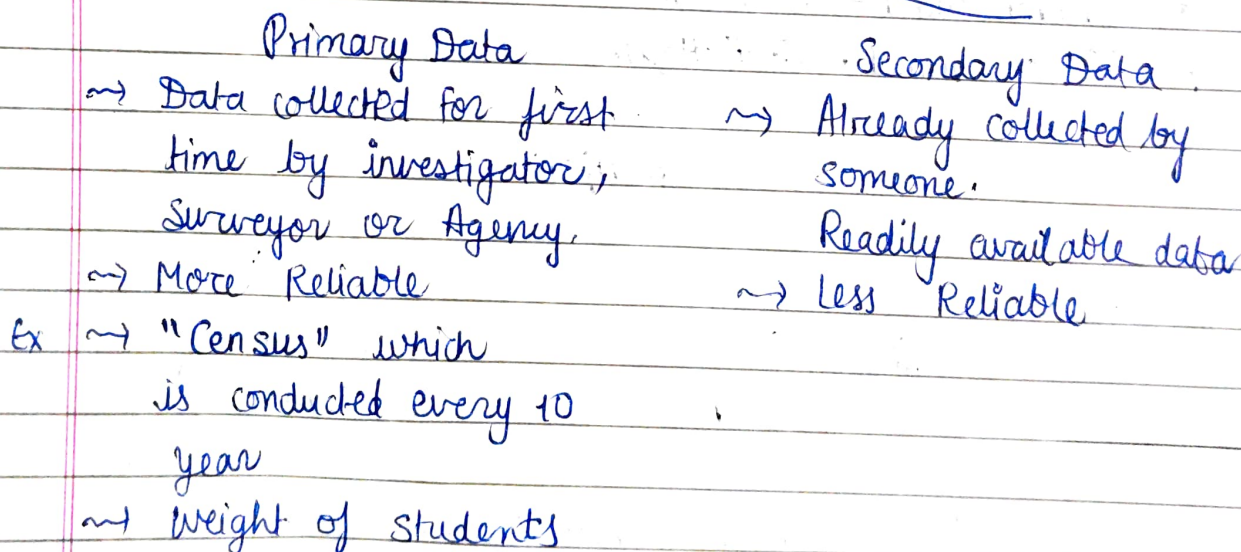
↑ quantitative

Data → Any information about some particular characteristics is known as data. (Numerical)

- Any information regarding to our study is known as data



Data (based on collection)



★ Methods of Collecting Primary Data.

(i) Interview Method.

→ Direct Personal Interview

Investigator meets the respondents directly & collects required information

Appropriate - In case of natural calamity like a super cyclone or an earthquake or an epidemic like plague

- Time consuming & costly
- Investigator must be trained.
- Not app. for large experiment.

→ Indirect Interview Method.

Used when there are some practical & technical problems in reaching the respondents directly.

Appropriate ^{then from the persons associated with the problems}

Road Accident, Bus Accident

Police Investigation, Marriage & time contact kisi orko krle h.

A third party is reached here.

→ Telephonic Interview

Information is ~~can be~~ gathered by the researcher himself by contacting interviewees over the phone.

★ Quickest Method & non-expensive way.

★ Non-Response is major de-merit ^{Maximum in this method}

★ less consistent but wide coverage

(ii) Mailed Questionnaire Method

drafted Method comprises of framing a well ~~defined~~ & soundly sequenced questionnaire covering all aspects of the problem under consideration. (Objective Question)

- ★ A wide area can be covered
- ★ Non-responses is maximum.

(iii) Observation Method (Direct)

By using necessary instruments or direct observation, investigator collects data.

- ★ like data on the height & weight of study
- ★ Best method for data collection
- ★ Time consuming
- ★ laborious
- ★ Covers only a small area.

(iv) Questionnaires filled & sent by enumerators. Used for larger enquiries.

Enumerators collect information directly by interviewing the persons having information.

- ★ Questions are explained & data is collected

Sources

Secondary Data



Published Sources

- * Government Reports like CSO
- * WHO, World Bank, IMF, ILO
- UNESCO (Recognised institutions)
- * Semi Govt sources (Quasi)
- ISI, ICAR, NCERT etc.
- * Internet/ TV/ etc (Newspaper)

Unpublished Sources

→ Reports collected from
Research Institutes,
Researchers

Collection of Data

Classification of Data

Scrutiny
of data

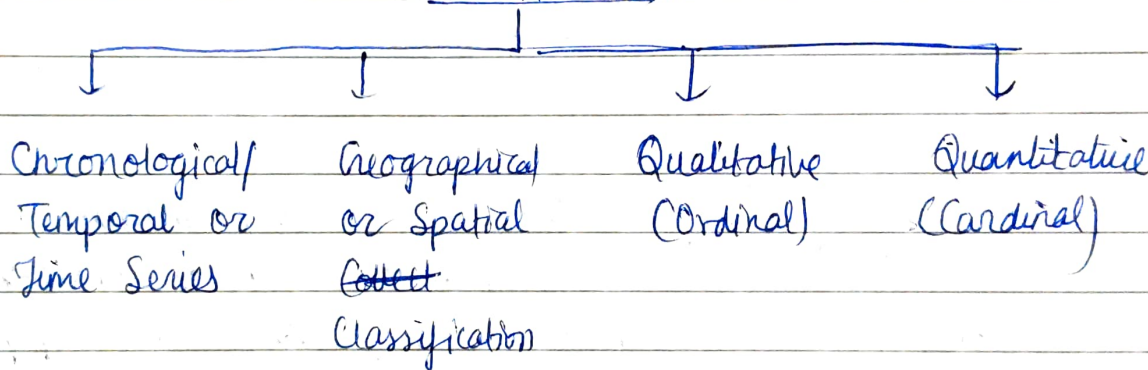
- * Scrutiny of Data → Checking of data whether it is accurate as well as consistence
- Errors may creep in while writing or copying the answer or on part of enumerator

Classification of Data

Arranging data on the basis of characteristics under consideration, into classes according to the similarities.

- ★ Puts the data in a neat, precise and condensed form which makes it easily understandable & interpretable.
- ★ Comparison possible thereby finding association or lack of it between series.
- ★ Statistical analysis is possible only for classified data.
- ★ Eliminates unnecessary data & make data readily understandable.

Classification



- | | | | |
|---|---|--|---|
| • On the basis of successive time points or intervals.
→ No. of students appeared for CA final for last twenty years | On the basis of place/ geography
→ Students appeared for exams from different states | On the basis of quality / Attributes
→ Nationality, Gender, Smoking habit | On the basis of variable
→ height, weight, profits, salaries etc |
|---|---|--|---|

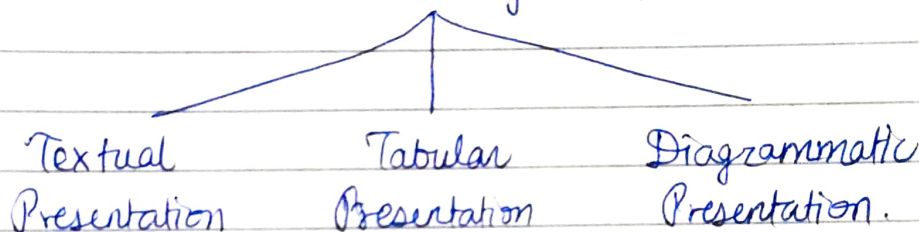
★ Data may be further classified as frequency data & non-frequency data.

The qualitative as well as quantitative data belong to the frequency group.

Time series data & geographical data belong to the non-frequency group.

3).

Presentation of Data



ci) Textual Presentation - Data is presented in form of paragraph or number of paragraphs

Example - Reports of Enquiry Commissions is usually made by textual representation.

Textual - [A survey shows 40% of total Popⁿ is coffee drinkers
30% of this popⁿ also drinks tea. out of non coffee drinkers 60% are male]

Merit → Merit lies in its simplicity & even a layman can present data by this method

→ Observations with exact magnitude can be presented with the help of this.

→ This method can be taken as the first step towards other modes of presentation.

Demerit → Dull, monotonous, comparison not possible

Tabular Method

Tabulation may be defined as systematic presentation of data with the help of a statistical table having a number of rows & columns.

★ Diagrammatic Representation of Data

Representation is made with the help of graphs & diagrams

→ It is the most attractive method to represent the data

Diagrammatic representation can be used for both the educated section & uneducated section.

Type of Diagrams

- | | |
|-----------------|----------------------|
| 1. Line Diagram | 4. Histogram |
| 2. Bar Diagram | 5. Ogive |
| 3. Pie-Chart | 6. Frequency Polygon |

Graphical Representation

c) Line Diagram - In this diagram, Data is represented by a line which is joined by plotting different points.

→ It is used for time series data.

→ Time is represented on x-axis and other variable is shown on y-axis.

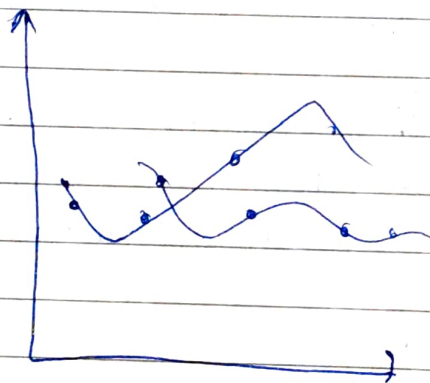
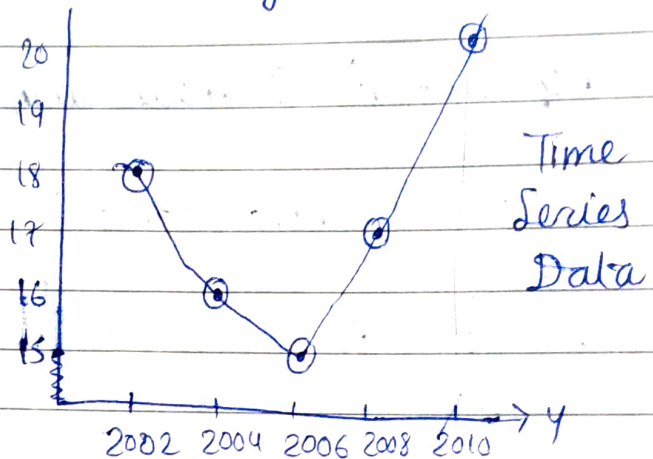
→ When we want to compare different type of series we use multiple line diagram.

→ When there is a lot of fluctuation between the value of 'y' variable we use 'log y' instead of 'y'.

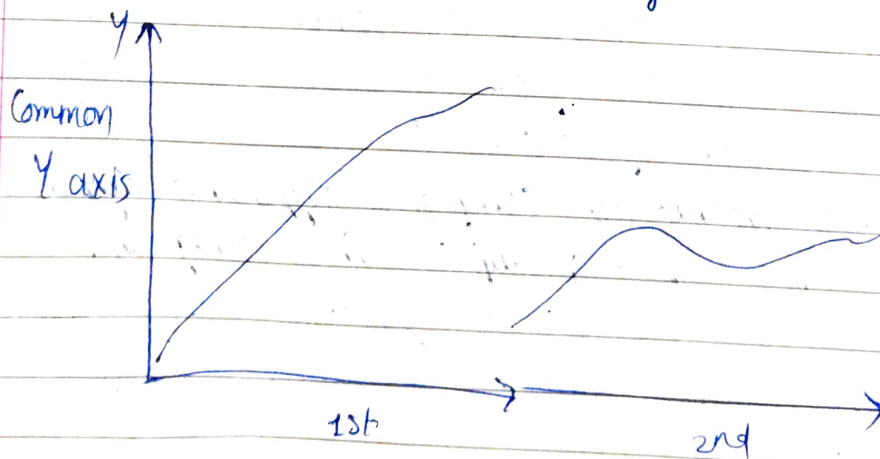
The chart is known as Logarithm chart or Ratio Chart.

It is also known as Historiogram

Year	Production
2002	18
2004	16
2006	15
2008	17
2010	20

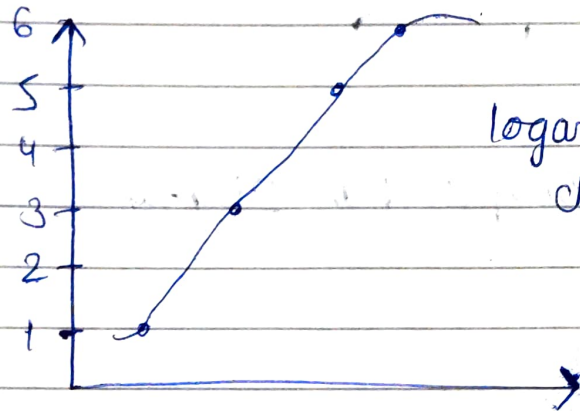


* Multiple Axis line Diagram is used when two things have different units
One has kilogram, 2nd Metre.



Time	Production	$\log y$
1999	10	1
2000	1000	3
2001	100000	5
2002	10,00,000	6
2003	50,00,000	6.6682

When fluctuations in y are huge, we cannot take true scale, so we use log.



logarithm / Ratio
chart chart

(2) Bar Graph:

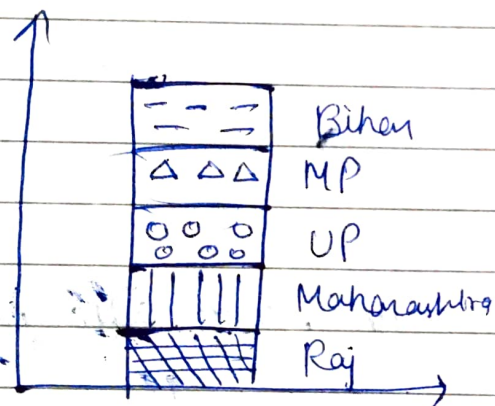
Bar - Rectangles of equal width, not adjacent, Equal gap



If one variable is shown in one diagram, it is known as Simple bar graph.



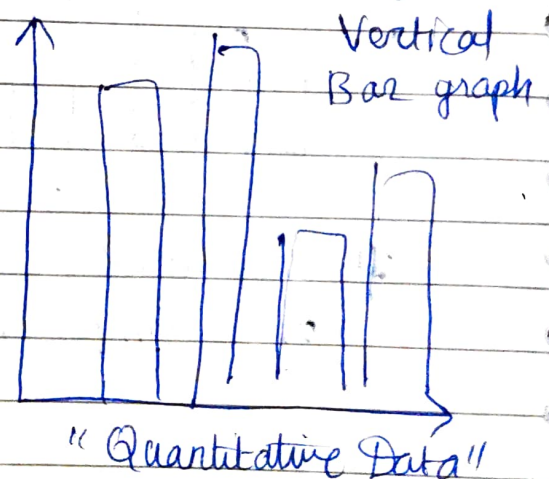
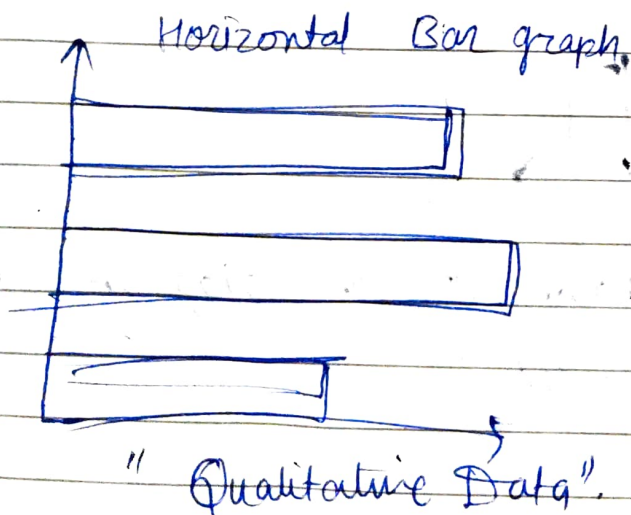
Multiple Bar Diagram



- Sub Divided Bar graph

One Bar is divided into different parts.

if converted in % then known as Percentage Bar graph.



(3) Pie-Chart

~ A pie chart is a Circular graph which represents the total value with its components.

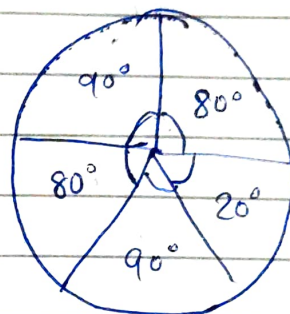
~ The Area of Circle represents the total value & the different sectors of circle represent the different parts.

~ Pie-chart is used for comparing different components of a variable & also relating the components on the whole

~ Total value of Central angle is 360°

~ It is area based diagram

~ In Pie diagram, Data are expressed in Percentage.



Components	Values	Angles	
A	6	18°	$\left(\frac{6}{120} \times 360^\circ \right)$
B	30	90°	$\left(\frac{30}{120} \times 360^\circ \right)$
C	48	144°	"
D	12	36°	"
E	<u>24</u>	72°	"
	120		

$$\text{Central Angles} = \frac{\text{Value}}{\text{Total value}} \times 360^\circ$$

★ Mostly used for 'Income-Expenditure' Bureau -

Variable

Constant

8

10

24

Variable →

x, y, z

"Height", "Temperature"

Weight, Income, Marks of Students?

Capable of changing its value

Variable

↓

Discrete
Variable

"Always Assumes
Integer value"

Number of students
in a class,

~~Marks~~ - Runs in a cricket

↓

Continuous
Variable

★ Temperature

36.1°C, 36.8°C

which assumes
any value in any
interval.

★ Height, Weight.

★ Marks of a student

Discrete variable

★ Attributes - Not Measurable / Diff from variable
It's a quality

Drinking habit of a person

Nationality of a person

Smoking habit

"Series"

↓

Individual
Series

Without any table
with comma.

No frequency.

8, 10, 7, 5, 8, 3

↓

Discrete
Series

Variable	freq.
5	2
8	1
10	3

↓

Continuous
Series

Class	frequency
0-10	5
10-20	6
20-30	7

Frequency → 8, 2, 3, 1, 8, 2, 3, 4

1 → freq. → 1

8 → freq. → 2

2 → freq. → 2

3 → freq. → 2

x	f	Tally Marks
1	1	I
2	4	IIII
3	1	I
4	2	II
5	3	III

Tally marks represent frequency

★ Cumulative frequency :- (C.F)

Class	frequency	CF
0-10	5	5
10-20	7	12
20-30	6	18
30-40	3	21

Continuous Series

Exclusive Series

Class	frequency
0-10	5
10-20	7
20-30	8
30-40	3

Includes
0-9 ←

Upper limit is excluded.

in every group

Overlapping Series

Inclusive Series

Class	frequency
10-19	5
20-29	7
30-39	3
40-49	2

Both the limits are

included but need

to be converted into

exclusive series

Non-overlapping series

Continuous Distribution

→ Classes (Divide)

→ frequency Distribute

Marks	No. of students
0-10	5
10-20	6
20-30	3
30-40	1

$$\text{Mid value} \rightarrow \frac{10 + 20}{2} = 15$$

$$\frac{UCL + LCL}{2}$$

Mid value $\Rightarrow$ Class marks

Inclusive Series $\longrightarrow$ Exclusive Series

$$\text{Adjusted factor} = \frac{g}{2} \quad (\text{gap})$$

$$\frac{1}{2} = 0.5$$

$$\begin{aligned} \text{LCB} &= \text{LCL} - \text{adjustment factor} \\ \text{UCB} &= \text{UCL} + \text{adjustment factor} \end{aligned} \quad \left| \frac{d}{2} \right.$$

$\hookrightarrow$ Upper class limit

$\hookrightarrow$ Upper class boundary

(Adjusted LCL & UCL of inclusive series)

Exclusive :- Class Interval = Upper Class Limit - lower Class Limit

Inclusive :- Class Interval = Upper Class Boundary - lower Class Boundary

* Relative Frequency

Class	Frequency	Rel. freq.
0-10	5	5/25
10-20	7	7/25
20-30	10	10/25
30-40	3	3/25
	<u>25</u>	

Relative frequency

$\Downarrow$

$$\frac{\text{frequency of class}}{\text{Total frequency}}$$

* Relative frequency always lies between '0' and '1'

* It is used for comparison between two or more series

- It is a Ratio of frequency of class to total frequency

Frequency Density

It is a ratio of frequency of class to class length

$$\text{Frequency Density} = \frac{\text{frequency of class}}{\text{Class length}}$$

Class length (Class width)

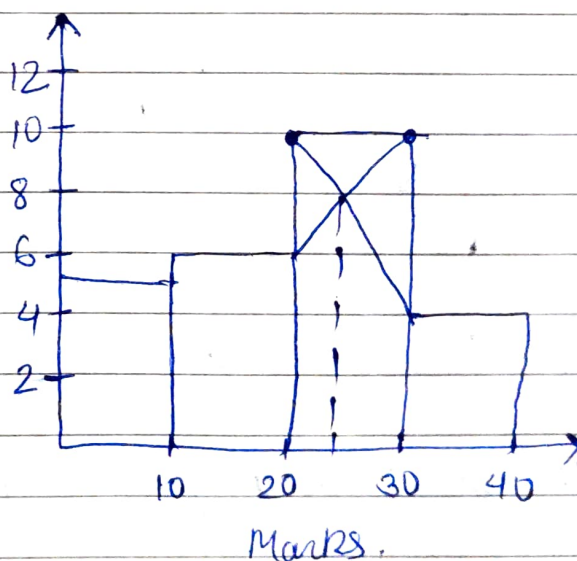
Class	frequency	frequency Density
10-15	5	$5/5 = 1$
15-18	6	$6/3 = 2$
18-24	12	$12/6 = 2$
24-30	6	$6/6 = 1$

★ Used in Histogram when Class intervals are not same.

Graphical Representation of Frequency Distribution.

Histogram

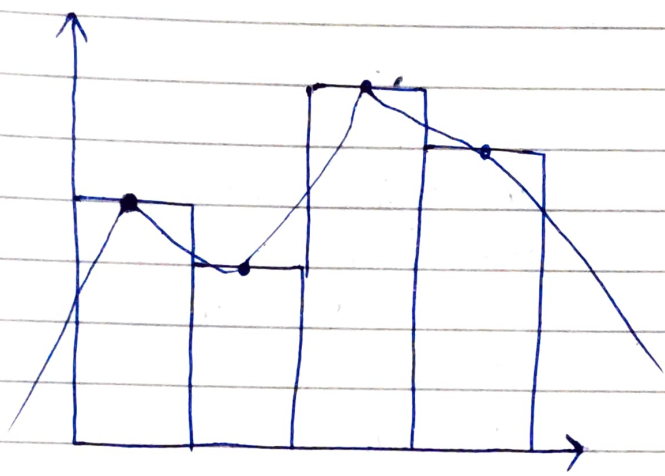
Marks	No. of students
0-10	5
10-20	6
20-30	10
30-40	4



- (1) Convenient way to represent continuous frequency distr
- (2) $x \rightarrow$ class boundaries
 $y \rightarrow$ frequency

- (3) Mode can be calculated by Histogram
(4) It is an area graph.

(2) Frequency Polygon

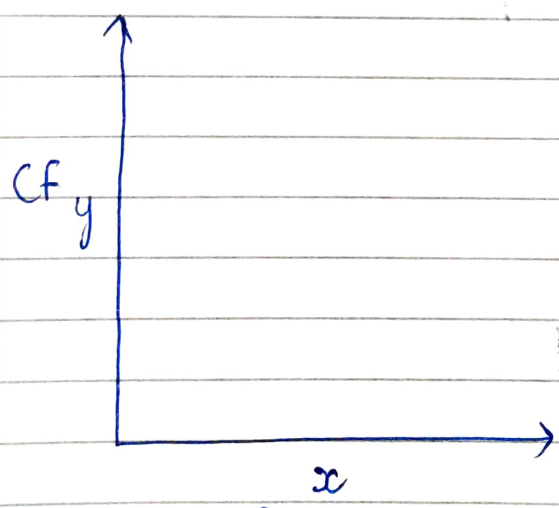


Frequency Polygon
"Mid Point of Histograms".

or
Simply by line diagram.

* It is a limiting of frequency curve.

(3) Ogive [Cumulative Frequency Curve]



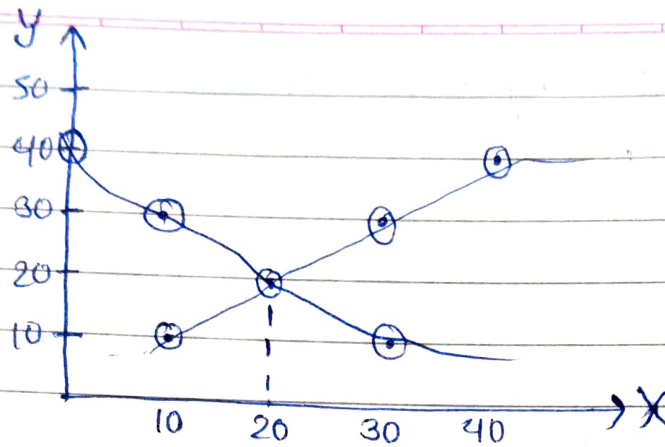
- It is a convenient way to represent cumulative frequency
- x axis - class boundaries
- y axis - C.F.

less than type More than type

Marks Class	f	CF
0-10	10	10
10-20	15	25
20-30	12	37
30-40	8	45

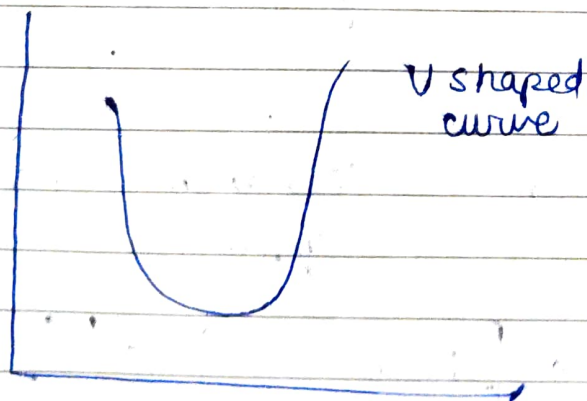
Marks	C.F.
less than 10	10
less than 20	25
less than 30	37
less than 40	45

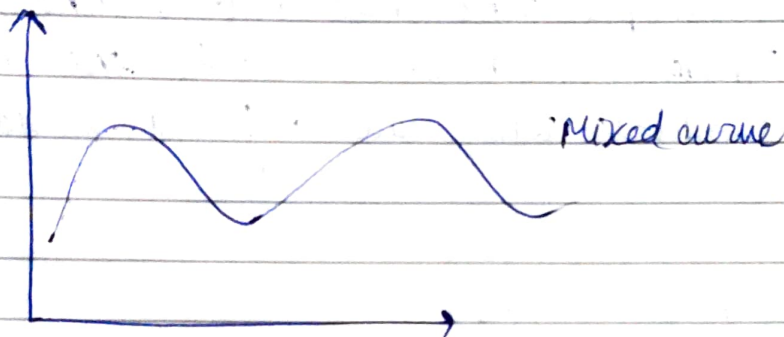
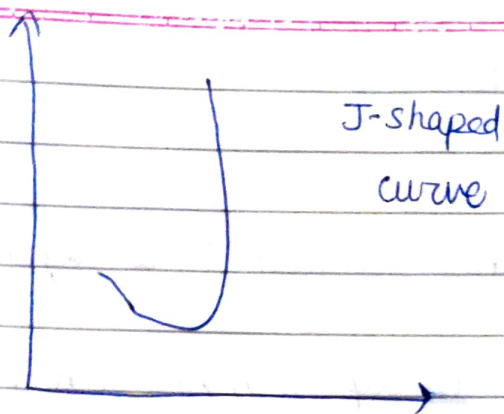
Marks	C.F.
More than 40	45
More than 30	35
More than 20	20
More than 10	8



- ★ The point where both ogive intersects is Median of that distribution.
- ★ Median, quartiles & other partition values can be obtained by ogive
- ★ less than type → Median

★ Frequency Curve





→ Dimensions of graph
(Length, Breadth, Height)

One Dimensional Graph :- Line "Diagram"
Bar Diagram.

Two Dimensional Graph :- "Pie Chart"
"Pie chart"
Histogram

Three Dimensional :- Square Diagram
Cubical, Cuboidal,
cylindrical graph

* Number of classes = $\frac{\text{Range}}{\text{Class Interval}}$

Find number of class - (Class size = 15)

125, 140, 128, 108, 96, 149, 136, 112, 84, 123, 130
120, 103, 89, 65, 103, 145, 97, 102, 87, 67, 78, 98,
126

$$\text{No. of classes} = \frac{149 - 65}{15} = \frac{84}{15} = 5.6$$

6