CHAPTER 13

6 MARKS

STATISTICAL DESCRIPTION OF DATA

BY : SHIVANI SHARMA

DEFINITION OF STATISTICS

SINGULAR SENSE

- scientific method that is employed for collecting, analysing and presenting data to draw statistical inferences

PLURAL SENSE

- data qualitative as well as quantitative, that are collected, usually with a view of having statistical analysis.

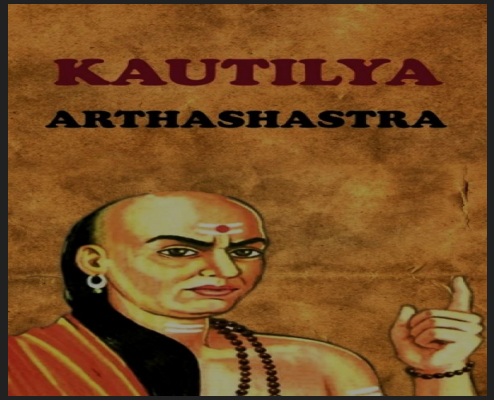
ORIGIN OF WORD – STATISTICS

Language	Word
LATIN	STATUS
ITALIAN	STATISTA
GERMAN	STATISTIK
FRENCH	STATISTIQUE

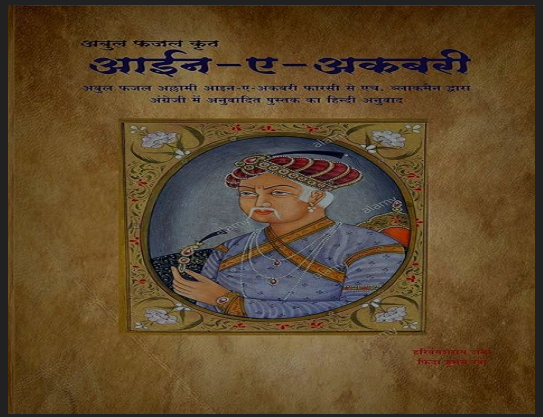
HISTORY OF STATISTICS

Kautilya's *'Arthashastra'* has record of births and deaths

during Chandragupta's reign in the fourth century B.C.



During the reign of Akbar in the sixteenth century A.D. We find statistical records on agriculture in *Ain-i-Akbari* written by Abu Fazl.



Referring to Egypt, the *first census* was conducted by the Pharaoh during 300 B.C. to 2000 B.C.

APPLICATIONS OF STATISTICS

- Economics
- Business Management
- Commerce and Industry

LIMITATIONS OF STATISTICS

I. Statistics deals with the aggregates and not individual data.

II. Statistics is concerned with quantitative data. However, qualitative data also can be converted to quantitative data by providing a numerical description to the corresponding qualitative data.

III. Future projections of sales, production, price and quantity etc. are possible under a specific set of conditions. If any of these conditions is violated, projections are likely to be inaccurate.

IV. Sampling based conclusions are used , improper sampling leads to improper results .

Quantitative information shown as number

DATA

PRIMARY

The data which are collected for the first time by an investigator or agency

SECONDARY

collected data used by a different person or agency.

Variable is a measurable data

VARIABLE

DISCRETE VARIABLE

- When a variable assumes a finite or a countably infinite number of isolated values, it is known as a discrete variable.
- **EXAMPLE :** Number of petals in a flower, the number of road accidents in locality

CONTINUOUS VARIABLE

- When a variable assumes any value from a given interval.
- **EXAMPLE :** height, weight

DISCRETE VARIABLE : Salary of a person

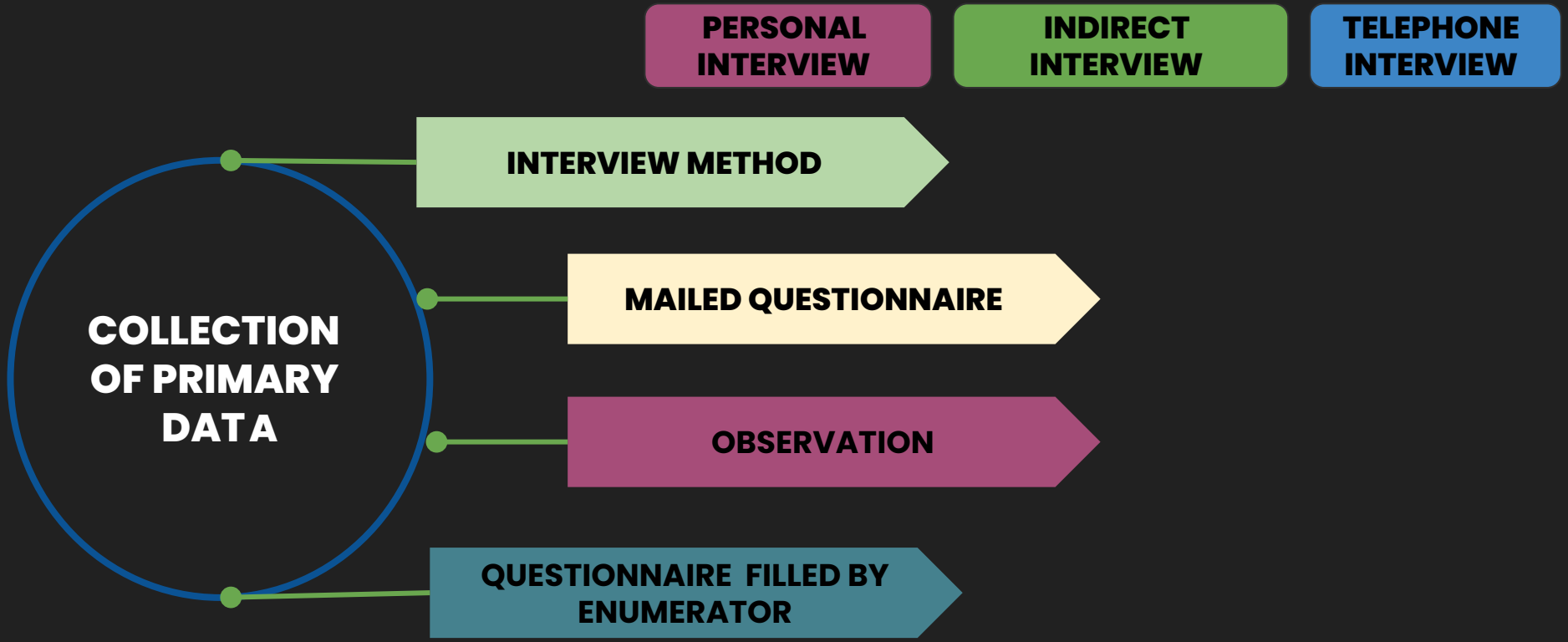
(Personal point of view)

CONTINUOUS VARIABLE : Turnover of a company

(Commercial point of view)

ATTRIBUTE

- A qualitative characteristic is known as an attribute.
- The gender of a baby, the nationality of a person, the colour of a flower etc. are examples of attributes.


COLLECTION OF PRIMARY DATA
INTERVIEW METHOD
PERSONAL INTERVIEW
INDIRECT INTERVIEW
TELEPHONE INTERVIEW
MAILED QUESTIONNAIRE
OBSERVATION
QUESTIONNAIRE FILLED BY ENUMERATOR

INTERVIEW METHOD

PERSONAL INTERVIEW METHOD

The investigator meets the respondents directly and collects the required information .

Highly accurate

EXAMPLE : *natural calamity like a super cyclone or an earthquake or an epidemic like plague,*

INDIRECT INTERVIEW METHOD

- When reaching respondent is difficult, data is collected by contacting associated persons .
- Highly accurate , low coverage
- **EXAMPLE :** *rail accident*

TELEPHONE INTERVIEW METHOD

Data is collected over phone

Quick and non -expensive method

Non-responses is maximum

Low accuracy

High coverage

MAILED QUESTIONNAIRE METHOD

- In this method well-drafted and soundly-sequenced questionnaire covering all the important aspects of the data requirement is sent to the respondents for filling .

- Coverage is wide but amount of non responses will be maximum

OBSERVATION METHOD

- In this method data is collected by direct observation or using instrument .
- **EXAMPLE :** *data on the height and weight of a group of students.*
- more accurate
- time consuming,
- laborious
- covers only a small area.

QUESTIONNAIRE FILLED AND SENT BY ENUMERATORS

- Enumerator means a Person who directly interacts with respondent and fills the questionnaire.
- It is generally used in case of surveys and census.

SOURCES OF SECONDARY DATA

- International sources : WHO, ILO, IMF, World Bank etc.
- Government sources : Statistical Abstract by CSO,
- Private and quasi-government sources : ISI, ICAR, NCERT etc.
- Unpublished sources of various research institutes, researchers etc.

SCRUTINY OF DATA

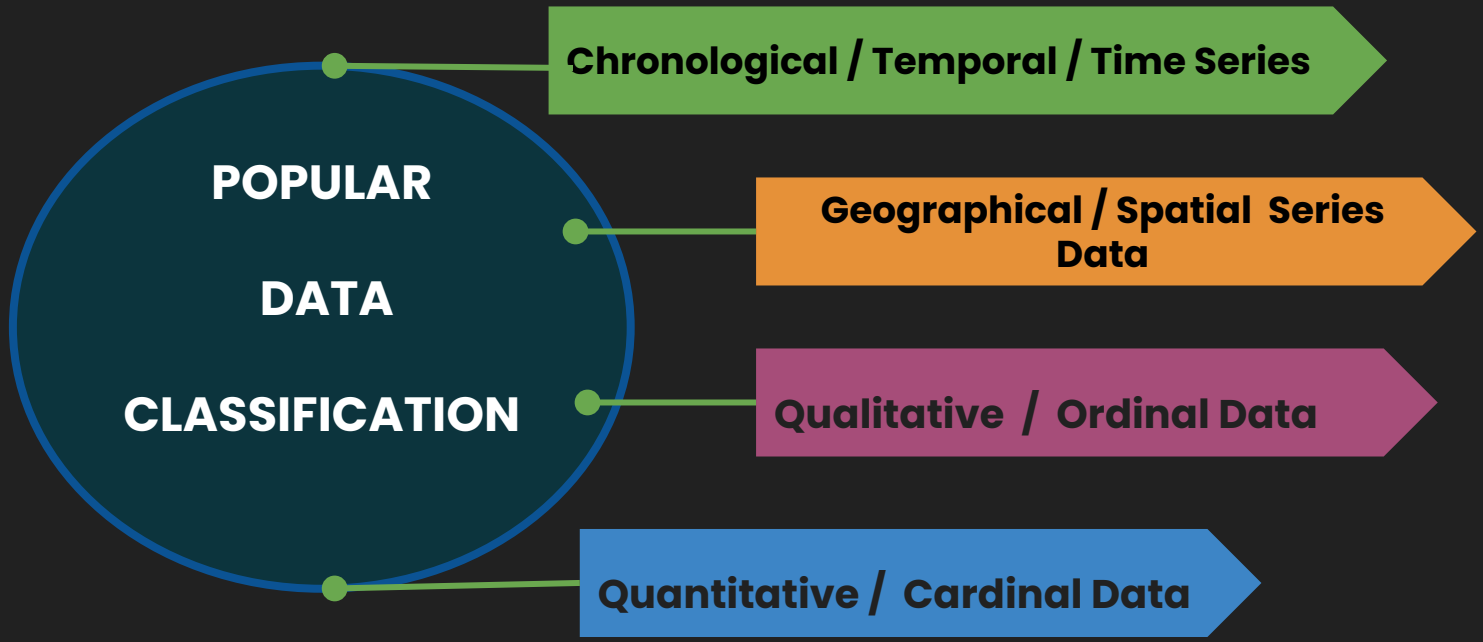
- Checking accuracy and consistency of data
- No hard and fast rules can be recommended for the scrutiny of data. One must apply his intelligence, patience and experience while scrutinising the given information.

INTERNAL CONSISTENCY

When two or more series of related data are given , we should check consistency among them

CLASSIFICATION OR ORGANISATION OF DATA

- It puts the data in a neat, precise and condensed form so that it is easily understood and interpreted.
- It makes comparison possible between various characteristics,
- Statistical analysis is possible only for the classified data.


POPULAR
DATA
CLASSIFICATION
Chronological / Temporal / Time Series
Geographical / Spatial Series Data
Qualitative / Ordinal Data
Quantitative / Cardinal Data

Chronological / Temporal / Time Series

When the data are classified in respect of successive time points or intervals, they are known as time series data.

EXAMPLE

The following example shows the population of India classified in terms of years.

Population of India (in crores)

Year	*Population (Crores)*
1951	35.7
1961	43.8
1971	54.6
1981	68.4
1991	81.8
2001	102.7
2011	121.0

Geographical / Spatial Series Data

Data arranged region wise are known as geographical data.

EXAMPLE

shows the yield of wheat in different countries

Yield of Wheat for Different Countries (2013)

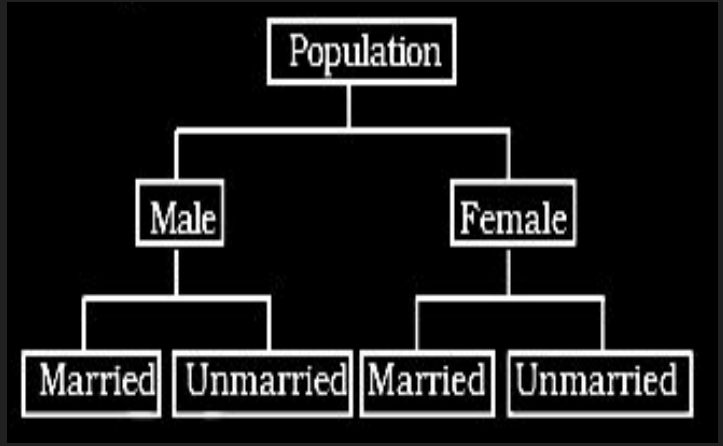
Country	*Yield of wheat (kg/hectare)*
Canada	3594
China	5055
France	7254
Germany	7998
India	3154
Pakistan	2787

QUALITATIVE / ORDINAL DATA

- Data classified in respect of an attribute are referred to as qualitative data. Data on nationality, gender, smoking habit of a group of individuals are examples of qualitative data.

EXAMPLE

In the following example, we find population of a country is grouped on the basis of the qualitative variable “gender”



Quantitative / Cardinal Data

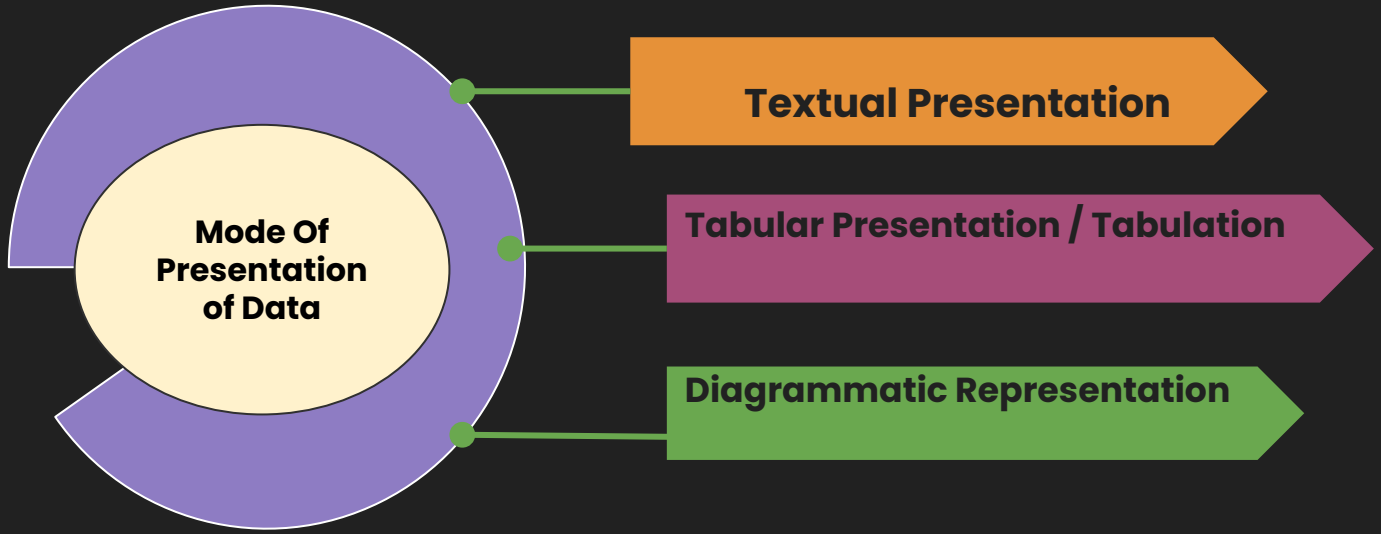
- when the data are classified in respect of a variable, say height, weight, profits, salaries , marks of students etc., they are known as quantitative data.

EXAMPLE

the quantitative classification of marks in mathematics

Frequency Distribution of Marks in Mathematics of 100 Students

Marks	*Frequency*
0–10	1
10–20	8
20–30	6
30–40	7
40–50	21
50–60	23
60–70	19
70–80	6
80–90	5
90–100	4
Total	*100*


Mode Of Presentation of Data
Textual Presentation
Tabular Presentation / Tabulation
Diagrammatic Representation

TEXTUAL PRESENTATION

- This method comprises presenting data with the help of a paragraph or a number of paragraphs.

- **EXAMPLE**
- 'In 2009, out of a total of five thousand workers of Roy Enamel Factory, four thousand and two hundred were members of a Trade Union. The number of female workers was twenty per cent of the total workers out of which thirty per cent were members of the Trade Union.

TEXTUAL PRESENTATION

MERITS

- Even a layman can present data by this method
- The observations with exact magnitude can be presented with the help of textual presentation.

DEMERITS

- It is dull, monotonous and comparison between different observations is not possible in this method.

TABULAR PRESENTATION / TABULATION

Tabulation may be defined as systematic presentation of data with the help of a statistical table .

MERITS

- It facilitates comparison between rows and columns.
- Complicated data can also be represented using tabulation.
- It is a must for diagrammatic representation.
- Without tabulation, statistical analysis of data is not possible.

TABULAR PRESENTATION / TABULATION

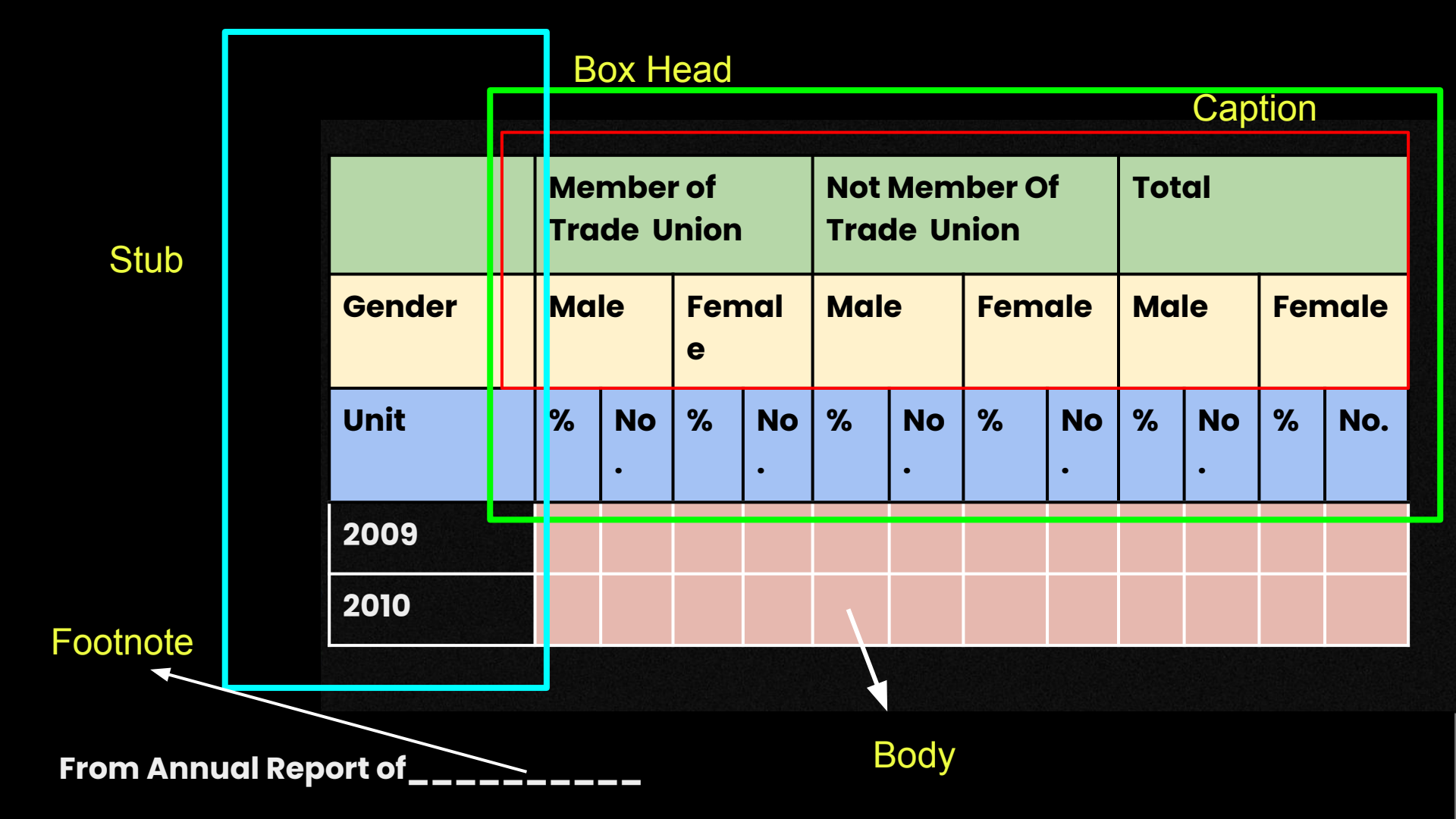
BOX HEAD : entire upper part of the table which includes columns and sub-column numbers, unit(s) of measurement along with caption.

CAPTION : the upper part of the table, describing the columns and sub-columns,

STUB : left part of the table providing the description of the rows.

BODY : main part of the table that contains the numerical figures.

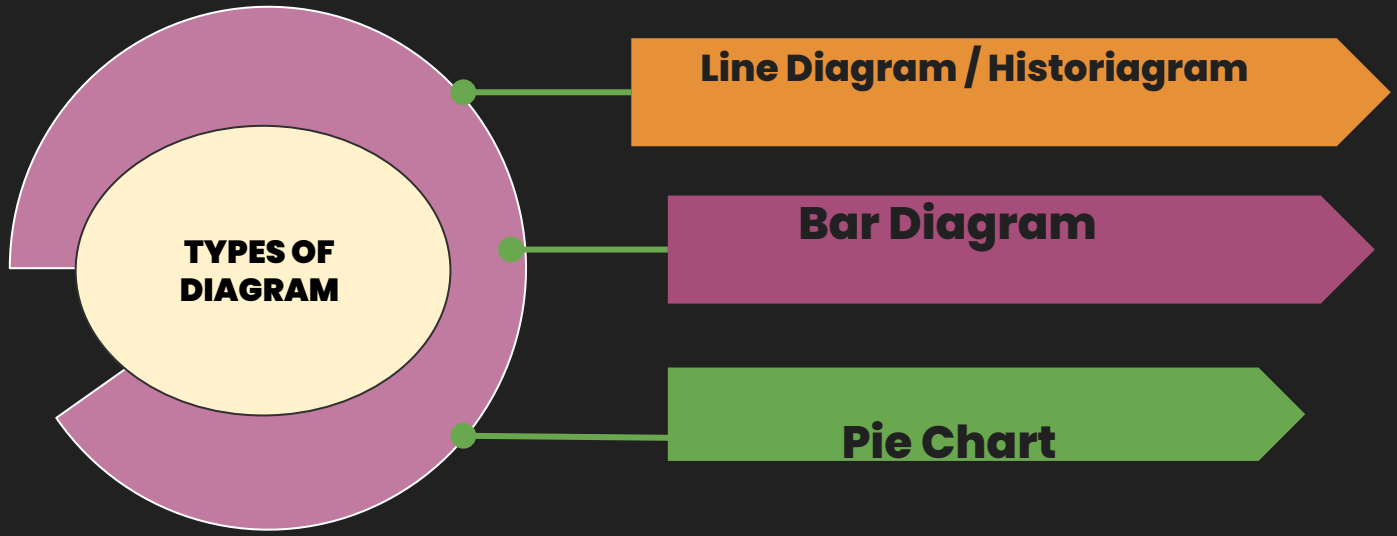
FOOTNOTE : source of the data at the bottom of table



	Member of Trade Union				Not Member Of Trade Union				Total			
Gender	Male		Female		Male		Female		Male		Female	
Unit	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.
2009												
2010												

DIAGRAMMATIC REPRESENTATION OF DATA

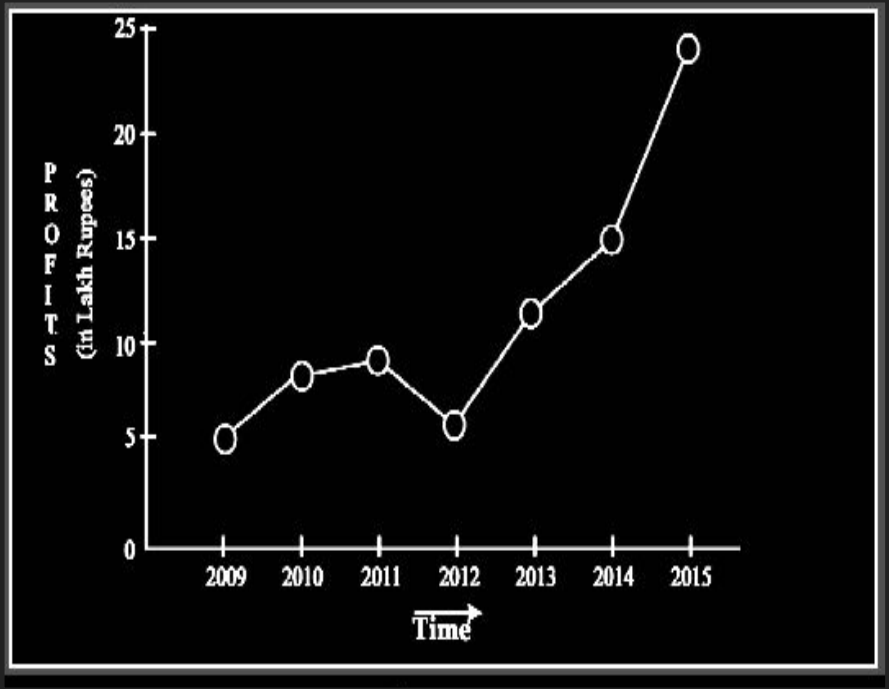
- An attractive representation of statistical data
- can be used for both the educated section and uneducated section of the society.
- Any hidden trend present in the given data can be noticed only in this mode of representation.
- Compared to tabulation, this is less accurate. So, if there is a priority for accuracy, we have to recommend tabulation.


TYPES OF DIAGRAM
Line Diagram / Historiagram
Bar Diagram
Pie Chart

Line Diagram or Historiagram

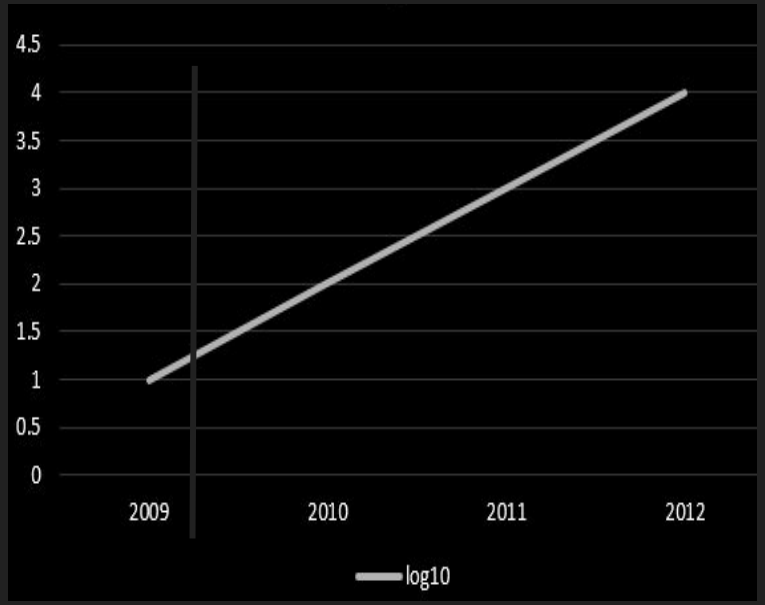
- Generally used for time series .
- For wide fluctuation : LOG CHART OR RATIO CHART is used
- For two or more series of same unit - MULTIPLE LINE CHART is used
- For two or more series of distinct unit - MULTIPLE AXIS CHART

LINE DIAGRAM



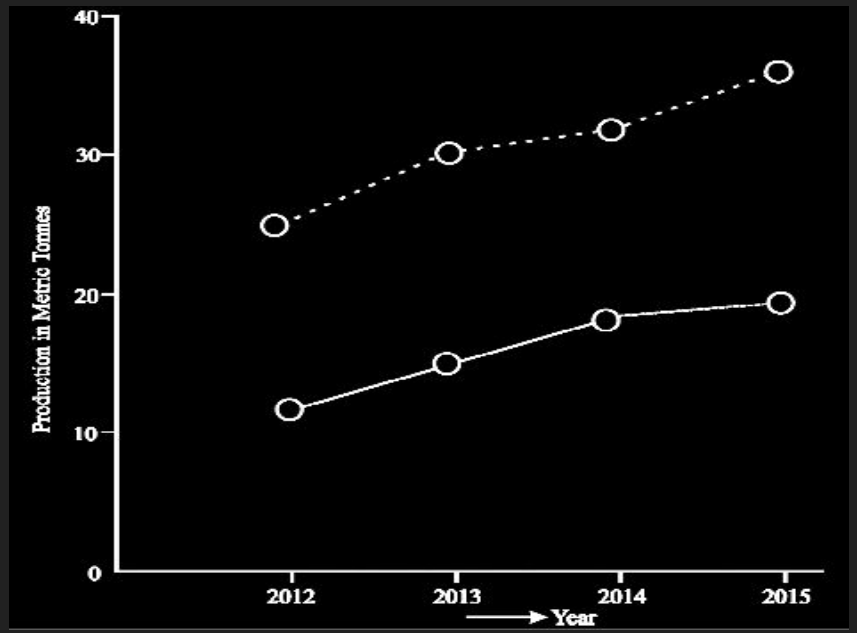
Year	Profit in Rs. Lakhs
2009	5
2010	8
2011	9
2012	6
2013	12
2014	15
2015	24

LOG CHART / RATIO CHART



Year x	Profit in Rs. Lakhs y	$\log_{10} y$
2009	10	1
2010	100	2
2011	1000	3
2012	10000	4

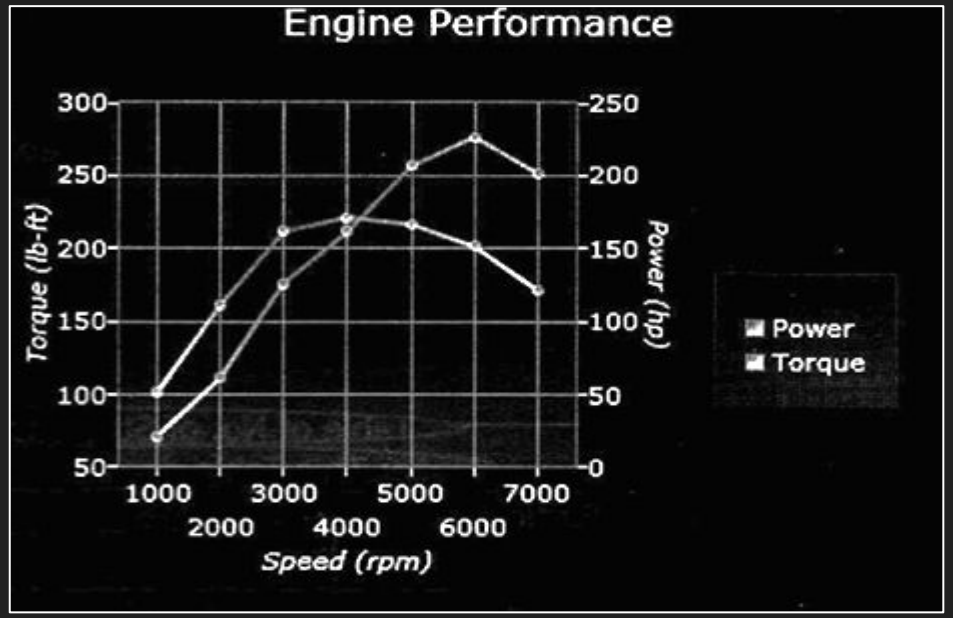
MULTIPLE LINE CHART



Year	Production in metric tones	
	Wheat	Rice
2012	12	25
2013	15	30
2014	18	32
2015	19	36

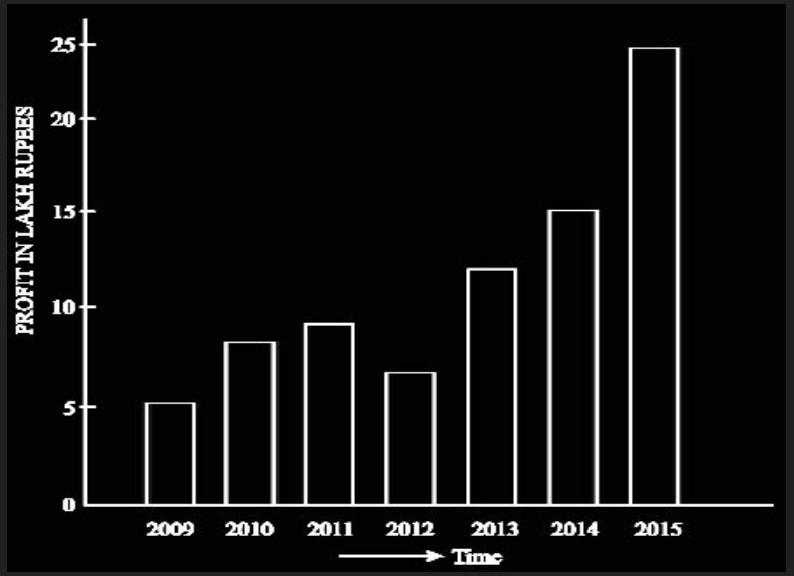
Dotted line represent production of rice and continuous line that of wheat

MULTIPLE AXIS CHART



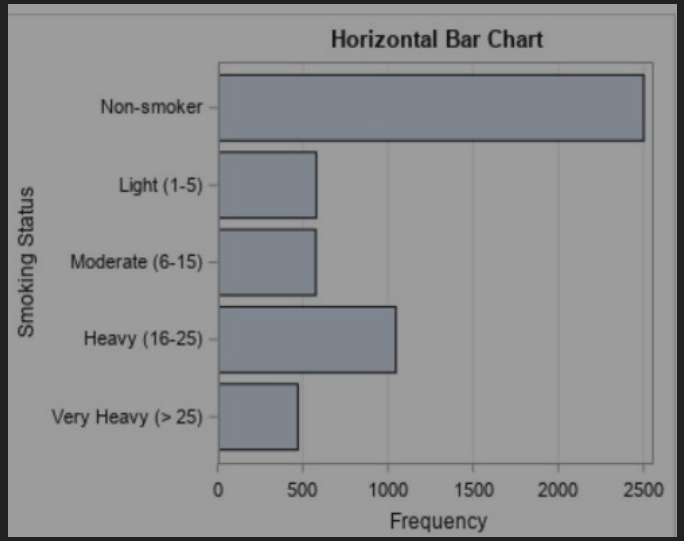
BAR DIAGRAM

Bars i.e. rectangles of equal width and usually of varying lengths drawn either horizontally or vertically.



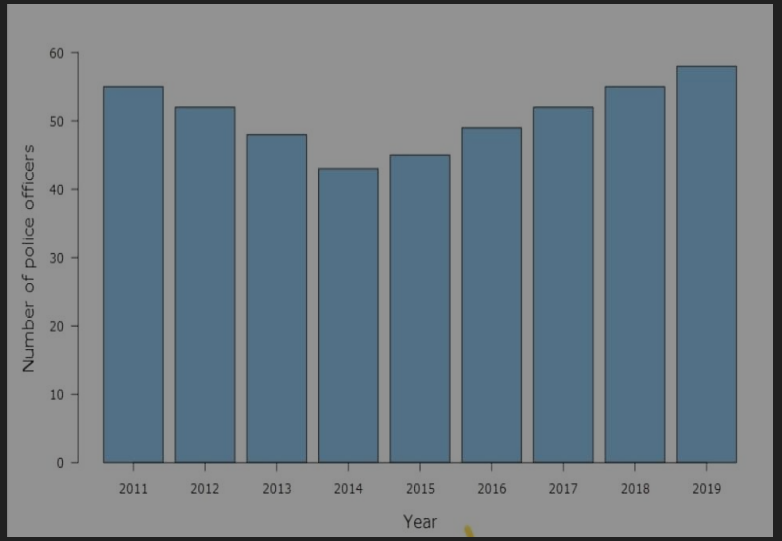
Year	Profit in Rs. Lakhs
2009	5
2010	8
2011	9
2012	6
2013	12
2014	15
2015	24

HORIZONTAL



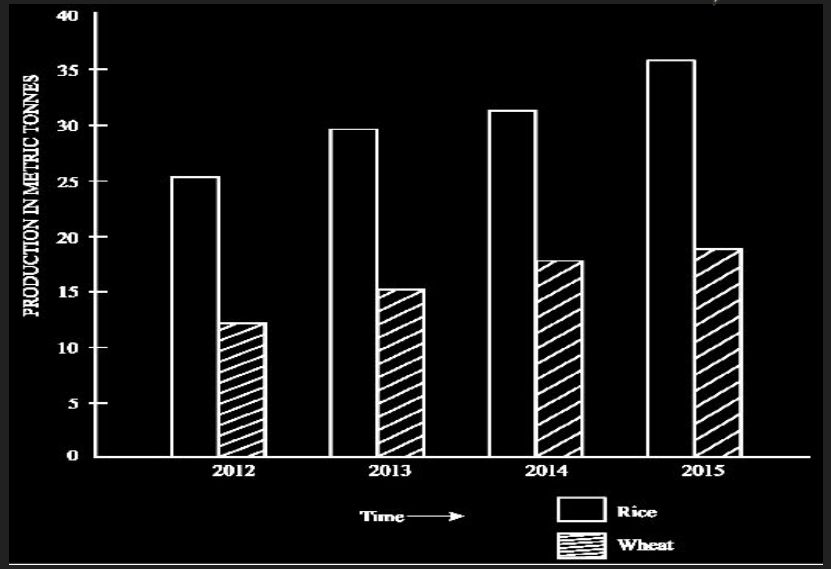
Qualitative data or Data varying over space (Geography)

VERTICAL



Quantitative data or Time series data

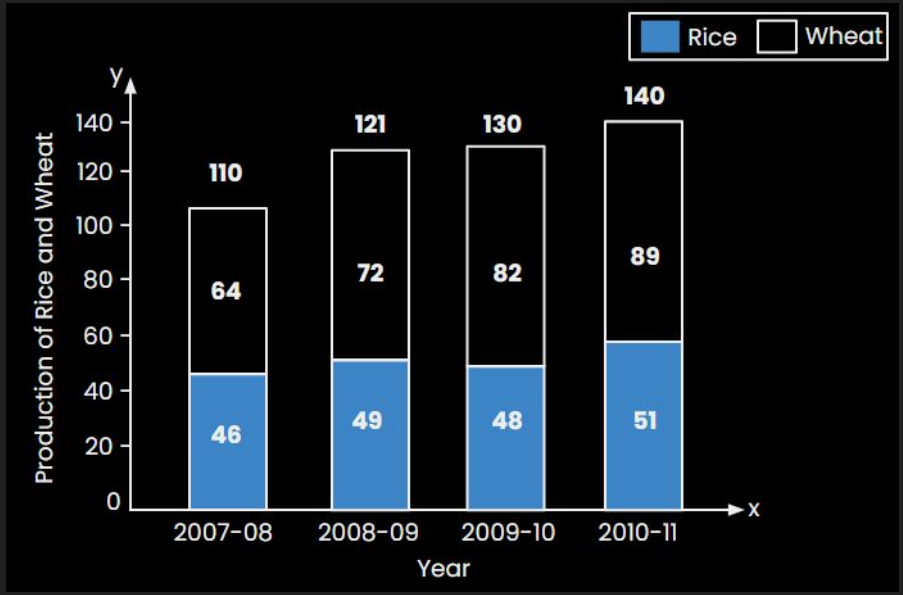
MULTIPLE / GROUPED BAR DIAGRAM



Year	Production in metric tones	
	Wheat	Rice
2012	12	25
2013	15	30
2014	18	32
2015	19	36

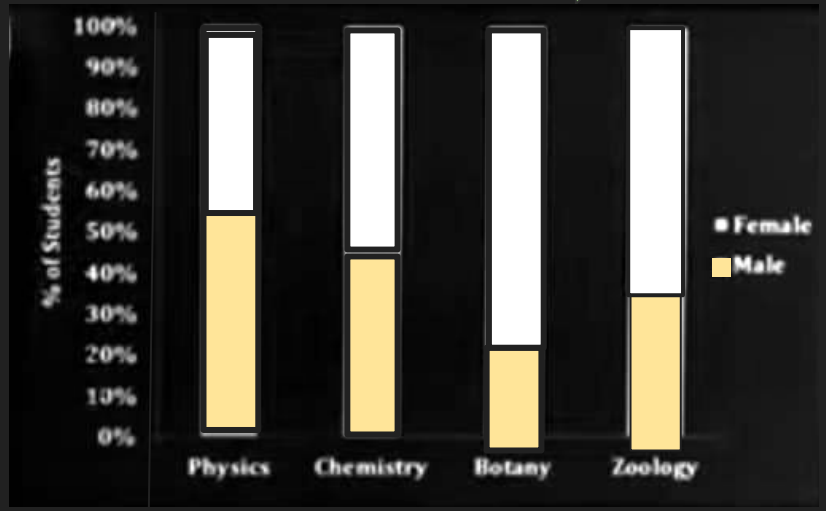
We consider Multiple or Grouped Bar diagrams to compare related series.

COMPONENT BAR DIAGRAM



- **Component or sub-divided Bar diagrams are applied for representing data divided into a number of components.**

Percentage BAR DIAGRAM



- For relative comparison to whole , percentage bar diagrams or divided bar diagrams are used

Pie chart

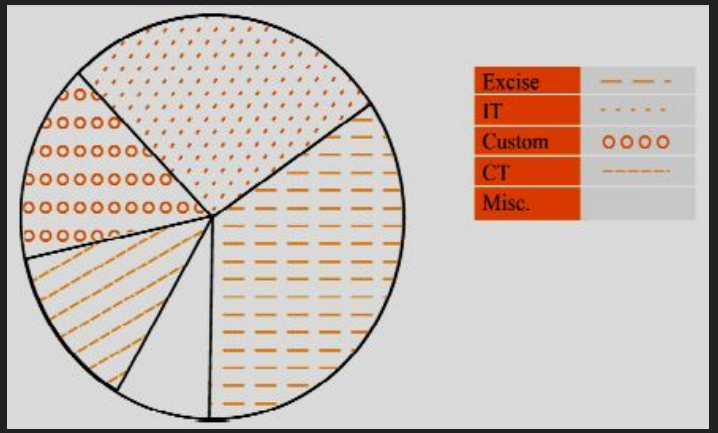
It is used for circular presentation of relative data

$$\text{Segment angle} = \frac{(\text{segment value} \times 360^0)}{(\text{total value})}$$

Example: Draw an appropriate diagram with a view to represent the following data :

Source	Revenue in millions of (₹)
Customs	80
Excise	190
Income Tax	160
Corporate Tax	75
Miscellaneous	35

Source (1)	Revenue in Million rupees (2)	Central angle $= \frac{(2)}{\text{Total of (2)}} \times 360^\circ$
Customs	80	$\frac{80}{540} \times 360^\circ = 53^\circ$ (approx.)
Excise	190	$\frac{190}{540} \times 360^\circ = 127^\circ$
Income Tax	160	$\frac{160}{540} \times 360^\circ = 107^\circ$
Corporate Tax	75	$\frac{75}{540} \times 360^\circ = 50^\circ$
Miscellaneous	35	$\frac{35}{540} \times 360^\circ = 23^\circ$
Total	540	360^0



Unit 1 Exercise
Set A

Que 1. Which of the following statements is false?

(a) Statistics is derived from the Latin word ‘Status’

(b) Statistics is derived from the Italian word ‘Statista’

(c) Statistics is derived from the French word ‘Statistik’

(d) None of these.

C

Que 2. Statistics is defined in terms of numerical data in the

(a) Singular sense

(b) Plural sense

(c) Either (a) or (b)

(d) Both (a) and (b).

b

Que 3. Statistics is applied in

(a) Economics

(b) Business management

(c) Commerce and industry

(d) All these.

d

Que 4. Statistics is concerned with

(a) Qualitative information

(b) Quantitative information

(c) (a) or (b)

(d) Both (a) and (b).

d

Que 5. An attribute is

(a) A qualitative characteristic

(b) A quantitative characteristic

(c) A measurable characteristic

(d) All these.

a

Que 6. Annual income of a person is

(a) An attribute

(b) A discrete variable

(c) A continuous variable

(d) (b) or (c).

b

Que 7. Marks of a student is an example of

(a) An attribute

(b) A discrete variable

(c) A continuous variable

(d) None of these.

b

Que. 8 Nationality of a student is

(a) An attribute

(b) A continuous variable

(c) A discrete variable

(d) (a) or (c).

a

Que 9 Drinking habit of a person is

(a) An attribute

(b) A variable

(c) A discrete variable

(d) A continuous variable.

a

Que 10. Age of a person is

(a) An attribute

(b) A discrete variable

(c) A continuous variable

(d) A variable.

C

Que 11. Data collected on religion from the census reports are

(a) Primary data

(b) Secondary data

(c) Sample data

(d) (a) or (b).

b

Que.12 The data collected on the height of a group of students after recording their heights with a measuring tape are

(a) Primary data

(b) Secondary data

(c) Discrete data

(d) Continuous data.

a

Que 13. The primary data are collected by

(a) Interview method

(b) Observation method

(c) Questionnaire method

(d) All these.

d

Que 14. The quickest method to collect primary data is

(a) Personal interview

(b) Indirect interview

(c) Telephone interview

(d) By observation.

C

Que 15. The best method to collect data, in case of a natural calamity, is

(a) Personal interview

(b) Indirect interview

(c) Questionnaire method

(d) Direct observation method.

a

Que 16. In case of a rail accident, the appropriate method of data collection is by

(a) Personal interview

(b) Direct interview

(c) Indirect interview

(d) All these.

C

Que 17. Which method of data collection covers the widest area?

(a) Telephone interview method

(b) Mailed questionnaire method

(c) Direct interview method

(d) All these.

b

Que 18. The amount of non-responses is maximum in

(a) Mailed questionnaire method

(b) Interview method

(c) Observation method

(d) All these.

a

Que 19. Some important sources of secondary data are

(a) International and Government sources

(b) International and primary sources

(c) Private and primary sources

(d) Government sources.

a

Que 20. Internal consistency of the collected data can be checked when

(a) Internal data are given

(b) External data are given

(c) Two or more series are given

(d) A number of related series are given.

d

Que 21. The accuracy and consistency of data can be verified by

(a) Internal checking

(b) External checking

(c) Scrutiny

(d) Both (a) and (b).

C

Que22. The mode of presentation of data are

(a) Textual, tabulation and diagrammatic

(b) Tabular, internal and external

(c) Textual, tabular and internal

(d) Tabular, textual and external.

a

Que23. The best method of presentation of data is

(a) Textual

(b) Tabular

(c) Diagrammatic

(d) (b) and (c).

b

Que24. The most attractive method of data presentation is

(a) Tabular

(b) Textual

(c) Diagrammatic

(d) (a) or (b).

C

Que 25. For tabulation, ‘caption’ is

(a) The upper part of the table

(b) The lower part of the table

(c) The main part of the table

(d) The upper part of a table that describes the column and sub-column.

d

Que 26. ‘Stub’ of a table is the

(a) Left part of the table describing the columns

(b) Right part of the table describing the columns

(c) Right part of the table describing the rows

(d) Left part of the table describing the rows.

d

Que 27. The entire upper part of a table is known as

(a) Caption

(b) Stub

(c) Box head

(d) Body.

C

Que28. The unit of measurement in tabulation is shown in

(a) Box head

(b) Body

(c) Caption

(d) Stub.

a

Que 29. In tabulation source of the data, if any, is shown in the

(a) Footnote

(b) Body

(c) Stub

(d) Caption.

a

Que 30. Which of the following statements is untrue for tabulation?

(a) Statistical analysis of data requires tabulation

(b) It facilitates comparison between rows and not columns

(c) Complicated data can be presented

(d) Diagrammatic representation of data requires tabulation.

b

Que 31. Hidden trend, if any, in the data can be noticed in

(a) Textual presentation

(b) Tabulation

(c) Diagrammatic representation

(d) All these.

C

Que. 32 Diagrammatic representation of data is done by

(a) Diagrams

(b) Charts

(c) Pictures

(d) All these.

d

Que33. The most accurate mode of data presentation is

(a) Diagrammatic method

(b) Tabulation

(c) Textual presentation

(d) None of these.

b

Que 34. The chart that uses logarithm of the variable is known as

(a) Line chart

(b) Ratio chart

(c) Multiple line chart

(d) Component line chart.

b

Que 35. Multiple line chart is applied for

(a) Showing multiple charts

(b) Two or more related time series when the variables are expressed in the same unit

(c) Two or more related time series when the variables are expressed in different unit

(d) Multiple variations in the time series.

b

Que 36. Multiple axis line chart is considered when

(a) There is more than one time series

(b) The units of the variables are different

(c) (a) or (b)

(d) (a) and (b).

d

Que 37. Horizontal bar diagram is used for

(a) Qualitative data

(b) Data varying over time

(c) Data varying over space

(d) (a) or (c).

d

Que 38. Vertical bar diagram is applicable when

(a) The data are qualitative

(b) The data are quantitative

(c) When the data vary over time

(d) (b) or (c).

d

Que 39. Divided bar chart is considered for

(a) Comparing different components of a variable

(b) The relation of different components to the table

(c) (a) or (b)

(d) (a) and (b).

d

Que 40. In order to compare two or more related series, we consider

(a) Multiple bar chart

(b) Grouped bar chart

(c) (a) or (b)

(d) (a) and (b).

C

Que 41 Pie-diagram is used for

(a) Comparing different components and their relation to the total

(b) Representing qualitative data in a circle

(c) Representing quantitative data in circle

(d) (b) or (c).

a

FREQUENCY And DISTRIBUTION

FREQUENCY : Number of times a particular observation is repeated.

FREQUENCY DISTRIBUTION TABLE : It is a table which contains observation or class intervals in one column and corresponding frequency in the other

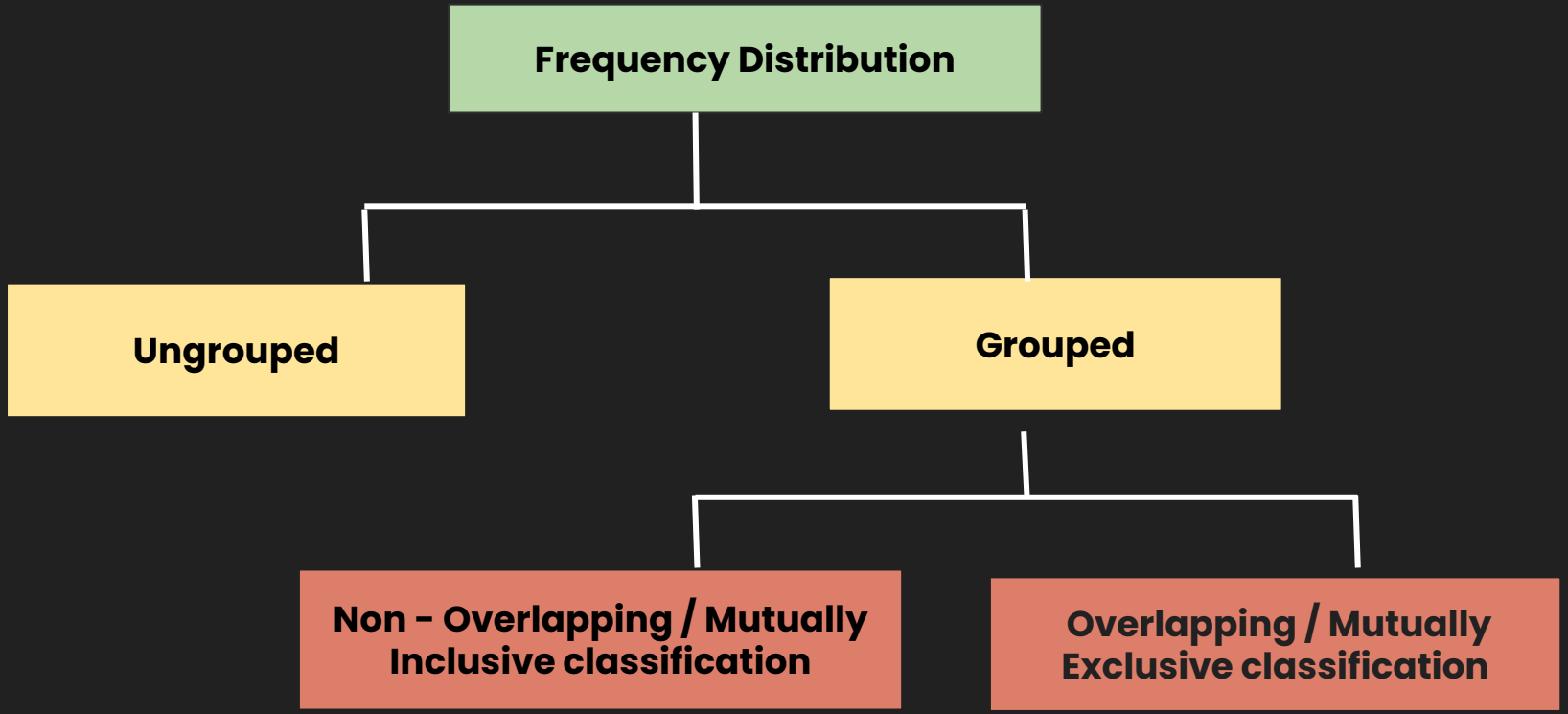
TYPES OF FREQUENCY DISTRIBUTION

Ungrouped / Simple Frequency Distribution

When there are limited number of distinct observations , frequency can be assigned to each one of them

Grouped Frequency distribution

When there are large number of observations , grouping is done among them , each group is called class interval and frequency is assigned to group and not individual value .


Frequency Distribution
Ungrouped
Grouped
Non - Overlapping / Mutually Inclusive classification
Overlapping / Mutually Exclusive classification

Class Limit

- For a class interval, the class limits may be defined as the minimum value and the maximum value the class interval
- Minimum value = lower class limit (LCL)
- Maximum value = upper class limit (UCL).

CLASS INTERVAL	FREQUENCY	LCL	UCL
44-48	3	44	48
49 - 53	4	49	53
54 -58	5	54	58

Non - Overlapping / Mutually Inclusive classification

CLASS INTERVAL	LCL	UCL
44-48	44	48
49 - 53	49	53
54 -58	54	58

- Includes UCL
- Usually applicable for discrete variable

Overlapping / Mutually Exclusive classification

CLASS INTERVAL	LCL	UCL
40 - 50	40	50
50 - 60	50	60
60 - 70	60	70

- Excludes UCL
- Usually applicable for continuous variable

CLASS BOUNDARY

OVERLAPPING / MUTUALLY EXCLUSIVE

CLASS INTERVAL	LCL	UCL	LCB	UCB
40 - 50	40	50	40	50
50 -60	50	60	50	60
60-70	60	70	60	70

Class limit = Class boundary

CLASS BOUNDARY

NON-OVERLAPPING / MUTUALLY INCLUSIVE

CLASS INTERVAL	LCL	UCL	LCB	UCB
44-48	44	48	43.5	48.5
49 - 53	49	53	48.5	53.5
54 -58	54	58	53.5	58.5

LCB = LCL - 0.5
UCB = UCL + 0.5

CLASS LENGTH

Class length = UCB - LCB

CLASS INTERVAL	LCL	UCL	LCB	UCB	CLASS LENGTH
44-48	44	48	43.5	48.5	5
49 - 53	49	53	48.5	53.5	5
54 -58	54	58	53.5	58.5	5

Mid-Point or Mid-Value or Class Mark

$$\text{mid-point} = \frac{\text{LCL} + \text{UCL}}{2} = \frac{\text{LCB} + \text{UCB}}{2}$$

CLASS INTERVAL	LCL	UCL	LCB	UCB	MID POINT
44-48	44	48	43.5	48.5	46
49 - 53	49	53	48.5	53.5	51
54 -58	54	58	53.5	58.5	56

Example: Following are the weights in kgs. of 36 BBA students of St. Xavier's College.

70	73	49	61	61	47	57	50	59
59	68	45	55	65	68	56	68	55
70	70	57	44	69	73	64	49	63
65	70	65	62	64	73	67	60	50

Construct a frequency distribution of weights, taking class length as 5.

Solution: We have, Range = Maximum weight – Minimum weight

= 73 kgs. – 44 kgs. = 29 kgs.

No. of class interval × class lengths = Range

No. of class interval × 5 = 29

No. of class interval = 29/5 = 6

(We always take the next integer as the number of class intervals so as to include both the minimum and maximum values).

Grouped Frequency Distribution

Weight in kg (Class Interval)	Tally marks	No. of Students (Frequency)
44-48	III	3
49-53	IIII	4
54-58	𝍸	5
59-63	𝍸 II	7
64-68	𝍸 IIII	9
69-73	𝍸 III	8
Total	–	36

Que 42. A frequency distribution

(a) Arranges observations in an increasing order

(b)Arranges observation in terms of a number of groups

(c) Relates to a measurable characteristics

(d) All of these

d

Que 43. The frequency distribution of a continuous variable is known as

(a) Grouped frequency distribution

(b) Simple frequency distribution

(c) (a) or (b)

(d) (a) and (b).

a

Que 44. The distribution of shares is an example of the frequency distribution of

(a) A discrete variable

(b) A continuous variable

(c) An attribute

(d) (a) or (c).

a

Que 45. The distribution of profits of a blue-chip company relates to

(a) Discrete variable

(b) Continuous variable

(c) Attributes

(d) (a) or (b).

b

Que 46. Mutually exclusive classification

(a) Excludes both the class limits

(b) Excludes the upper class limit but includes the lower class limit

(c) Includes the upper class limit but excludes the upper class limit

(d) Either (b) or (c).

b

Que 47. Mutually inclusive classification is usually meant for

(a) A discrete variable

(b) A continuous variable

(c) An attribute

(d) All these.

a

Que 48. Mutually exclusive classification is usually meant for

(a) A discrete variable

(b) A continuous variable

(c) An attribute

(d) Any of these.

b

Que 49. The LCB is

(a) An upper limit to LCL

(b) A lower limit to LCL

(c) (a) and (b)

(d) (a) or (b).

b

Que 50. The UCB is

(a) An upper limit to UCL

(b) A lower limit to LCL

(c) Both (a) and (b)

(d) (a) or (b).

a

Que 51. length of a class is

(a) The difference between the UCB and LCB of that class

(b) The difference between the UCL and LCL of that class

(c) (a) or (b)

(d) Both (a) and (b).

a

CUMULATIVE FREQUENCY

- These are of two types -

  Less than type cumulative frequency

  More than type cumulative frequency

- For a particular class boundary

  Less than type CF + More than type CF = Total frequency

Class	Frequency
0 – 10	4
10 – 20	8
20 –30	13
30 – 40	12
40 –50	6

Class Boundary	More than	Less than
0	43	0
10	39	4
20	31	12
30	18	25
40	6	37
50	0	43

Class	Frequency	More than (LCB)	Less than (UCB)
0 - 10	4	43	4
10 - 20	8	39	12
20 -30	13	31	25
30 - 40	12	18	37
40 -50	6	6	43

Frequency Density

$$\text{Frequency Density} = \frac{\text{Class Frequency}}{\text{Class Length of Class}}$$

Relative Frequency

$$\text{Relative frequency} = \frac{\text{Class Frequency}}{\text{Total Frequency}}$$

Relative frequencies add up to unity

Relative frequency for a particular class

Lies between 0 and 1

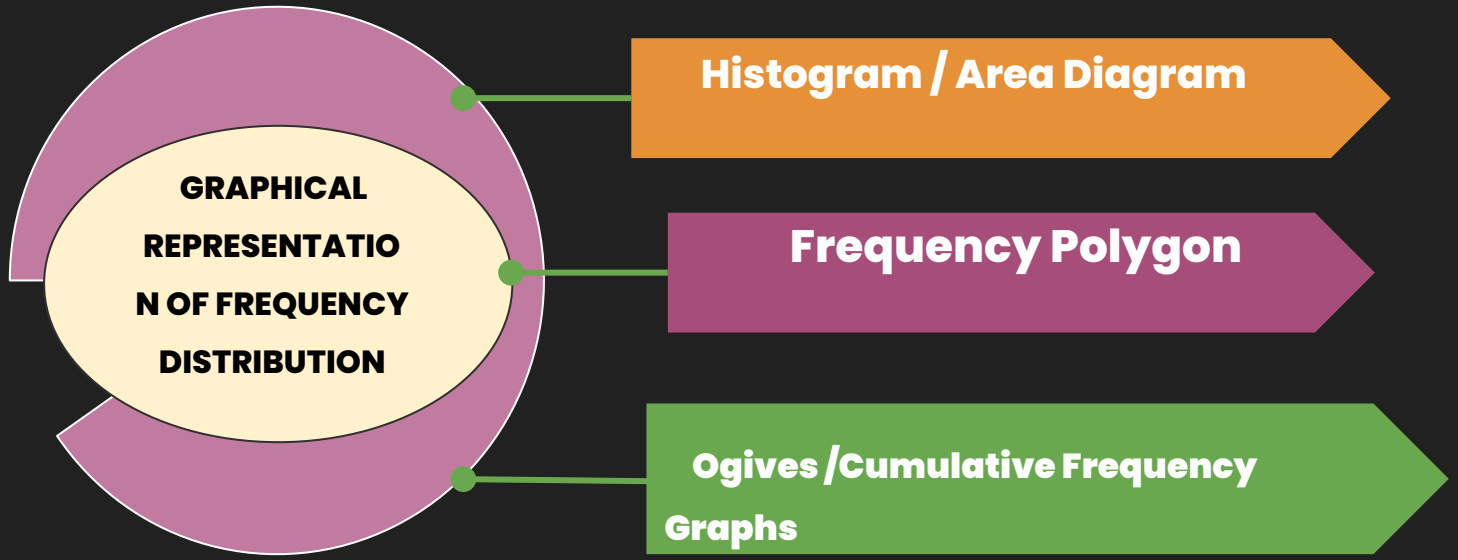
Percentage Frequency

$$\text{Percentage Frequency} = \frac{\text{Class Frequency}}{\text{Total Frequency}} \times 100$$

percentage frequencies add up to one hundred.

Class Interval	Frequency	Class Length	Frequency Density
44 - 48	3	5	3/5 = 0.6
49- 53	4	5	4/5 = 0.8
54- 58	5	5	
59 -63	7	5	
64 - 68	9	5	
69 -73	8	5	
Total	36		

Class Interval	Frequency	Relative Frequency	Percentage Frequency
44 - 48	3	**3/ 36 = 0.083**	**3/36 x 100 = 8.33 %**
49- 53	4	**4/36 = 0.111**	
54- 58	5	**5/36**	
59 -63	7	**7/36**	
64 - 68	9	**9/36**	
69 -73	8	**8/36**	
Total	36	1	100 %

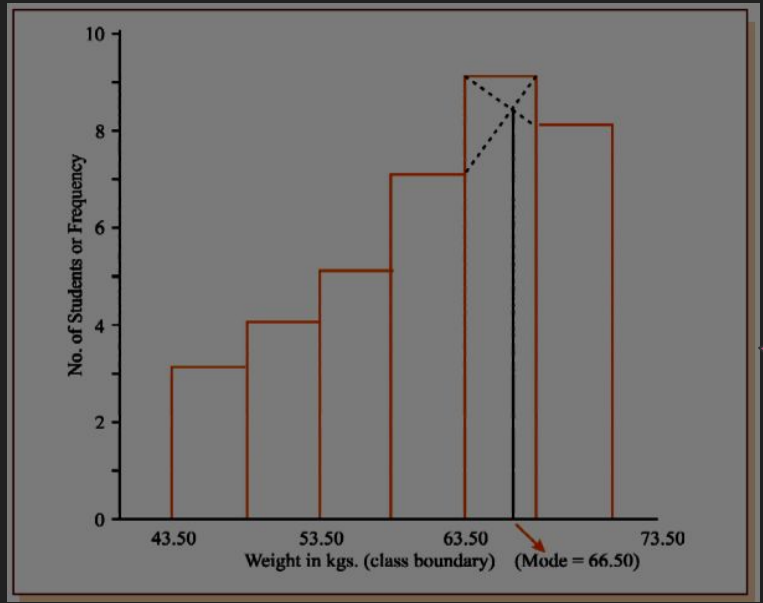

GRAPHICAL REPRESENTATIO N OF FREQUENCY DISTRIBUTION
Histogram / Area Diagram
Frequency Polygon
Ogives /Cumulative Frequency Graphs

HISTOGRAM / AREA DIAGRAM

- This is a very convenient way to represent a frequency distribution.
- Comparison between frequency of two different classes are possible
- It is used to calculate **MODE.**

HISTOGRAM / AREA DIAGRAM

Class Interval	Frequency	LCB	UCB
44 - 48	3	43.5	48.5
49- 53	4	48.5	53.5
54- 58	5	53.5	58.5
59 -63	7	58.5	63.5
64 - 68	9	63.5	68.5
69 -73	8	68.5	73.5
Total	36		



MODE

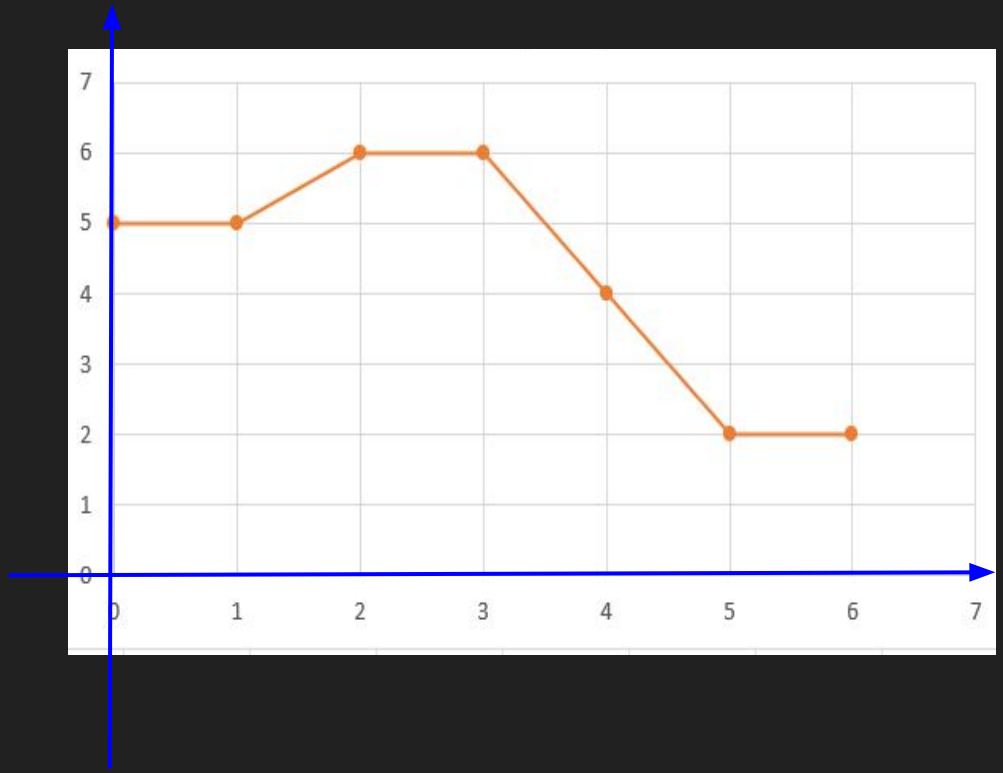
- Mode = 66.50 kgs.

FREQUENCY POLYGON

- Usually frequency polygon is meant for simple / Ungrouped frequency distribution.
- However, we also apply it for grouped frequency distribution provided the width of the class intervals remains the same.
- We can also obtain a frequency polygon starting with a histogram by adding the mid- points of the upper sides of the rectangles successively and then completing the figure by joining the two ends as before.

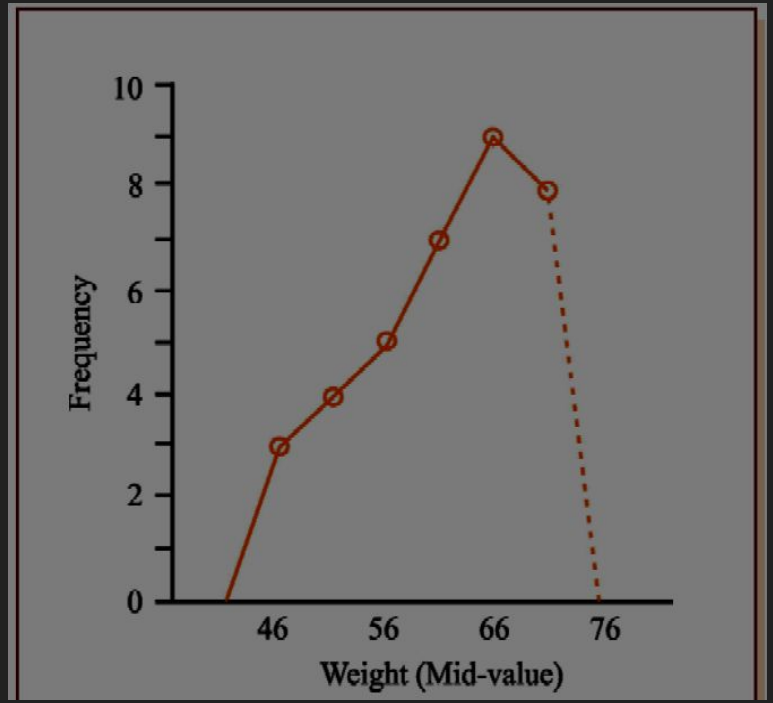
FREQUENCY POLYGON - UNGROUPED FREQUENCY DISTRIBUTION

Observation (x)	Frequency
0	5
1	5
2	6
3	6
4	4
5	2
6	2



FREQUENCY POLYGON – GROUPED FREQUENCY DISTRIBUTION

Mid-points	No. of Students (Frequency)
46	3
51	4
56	5
61	7
66	9
71	8



OGIVES / CUMULATIVE FREQUENCY GRAPH

By plotting cumulative frequency against the respective class boundary, we get ogives

TWO TYPES OF OGIVES

Less than type Ogives

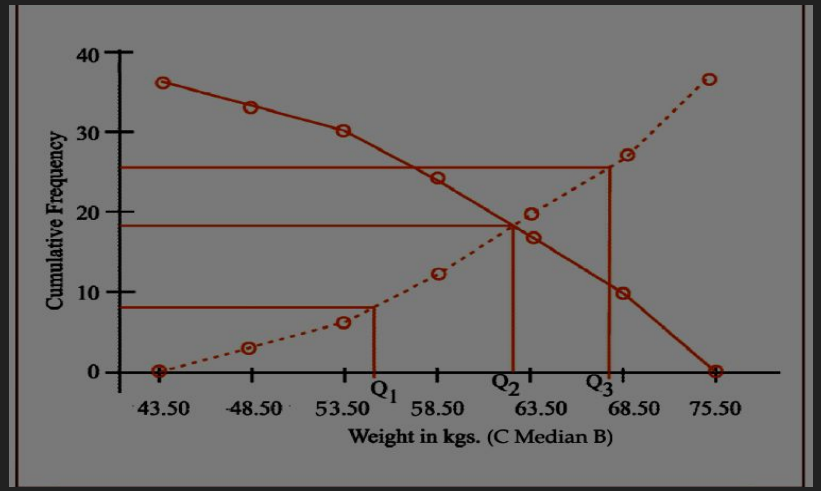
- less than type ogives, obtained by taking less than cumulative frequency on the vertical axis

More than type Ogives

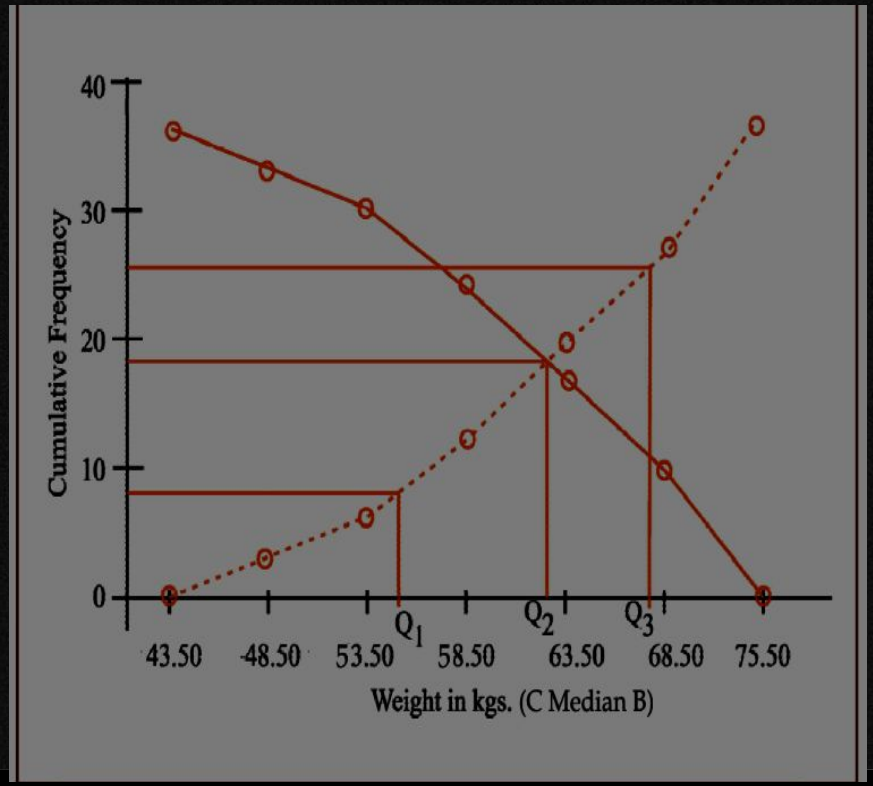
- more than type ogives by plotting more than type cumulative frequency on the vertical axis

OGIVES / CUMULATIVE FREQUENCY GRAPH

- Ogives may be considered for obtaining quartiles graphically.
- If a perpendicular is drawn from the point of intersection of the two ogives on the horizontal axis, then the x-value of this point gives us the value of median

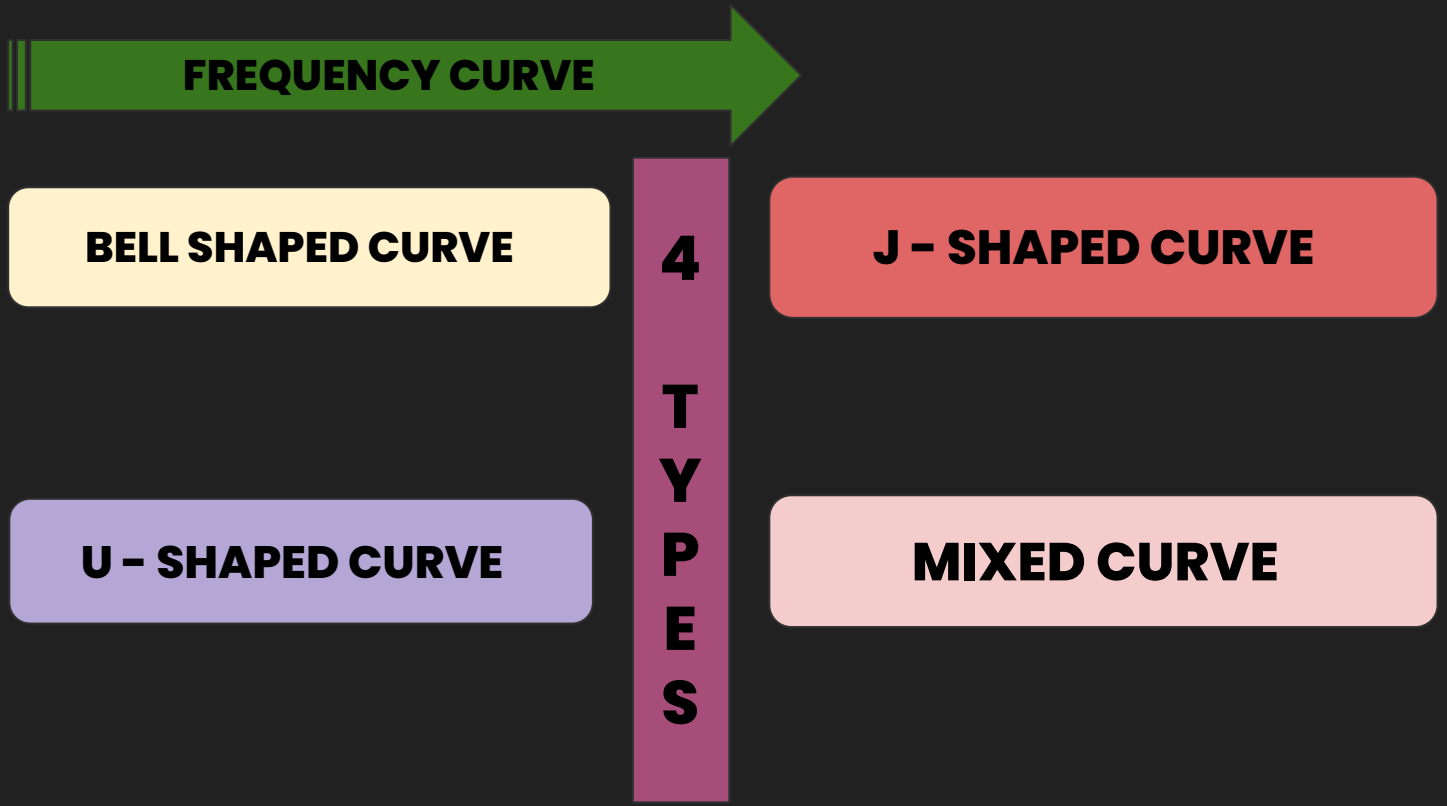


Class Interval	Frequency	CB	Less than type CF	More than type CF
44 - 48	3	43.5	**0**	**36**
49- 53	4	48.5	**3**	**33**
54- 58	5	53.5	**7**	**29**
59 -63	7	58.5	**12**	**24**
64 - 68	9	63.5	**19**	**17**
69 -73	8	68.5	**28**	**8**
		73.5	**36**	**0**



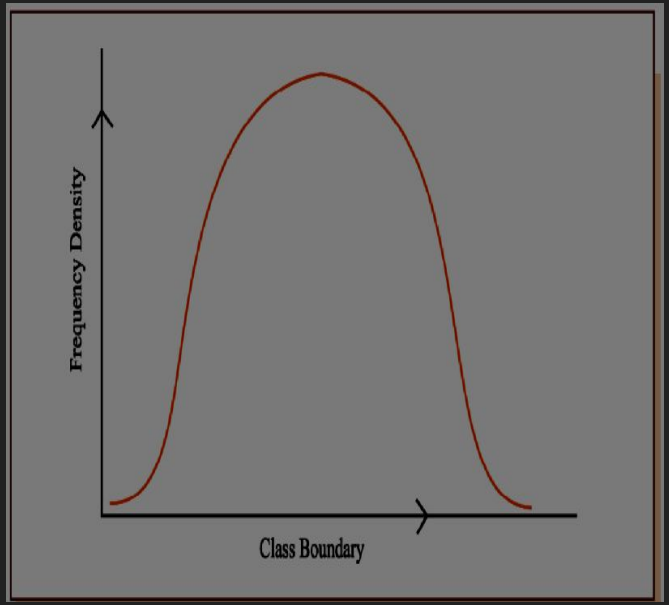
FREQUENCY CURVE

- It is a limiting form of a histogram or frequency polygon.
- The frequency curve for a distribution can be obtained by drawing a smooth and free hand curve through the mid-points of the upper sides of the rectangles forming the histogram.


FREQUENCY CURVE
BELL SHAPED CURVE
4
TYPES
J - SHAPED CURVE
U - SHAPED CURVE
MIXED CURVE

BELL SHAPED CURVE

- Most of the commonly used distributions provide bell-shaped curve, which, as suggested by the name, looks almost like a bell.

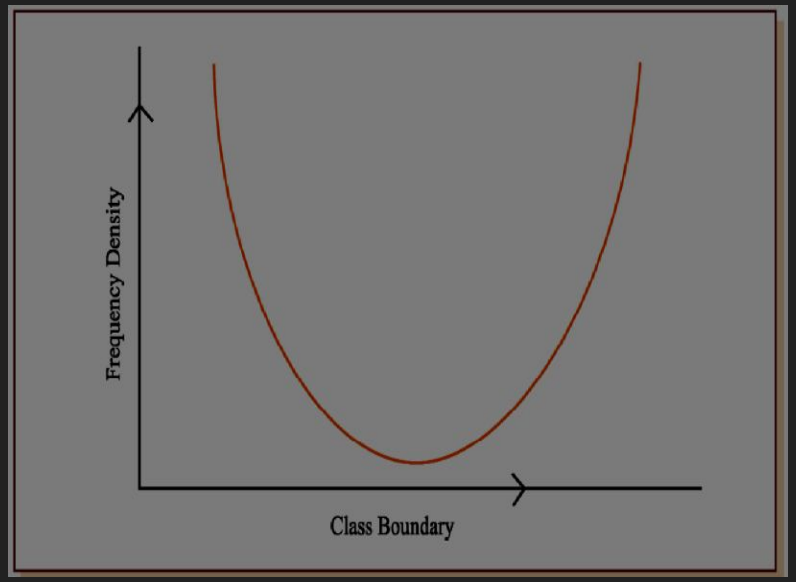


- The distribution of height, weight, mark, profit etc. usually belong to this category.

- On a bell-shaped curve , the frequency , starting from a rather low value , gradually reaches the maximum value , somewhere near the central part and then gradually decreases to reach its lowest value at the other extremity .

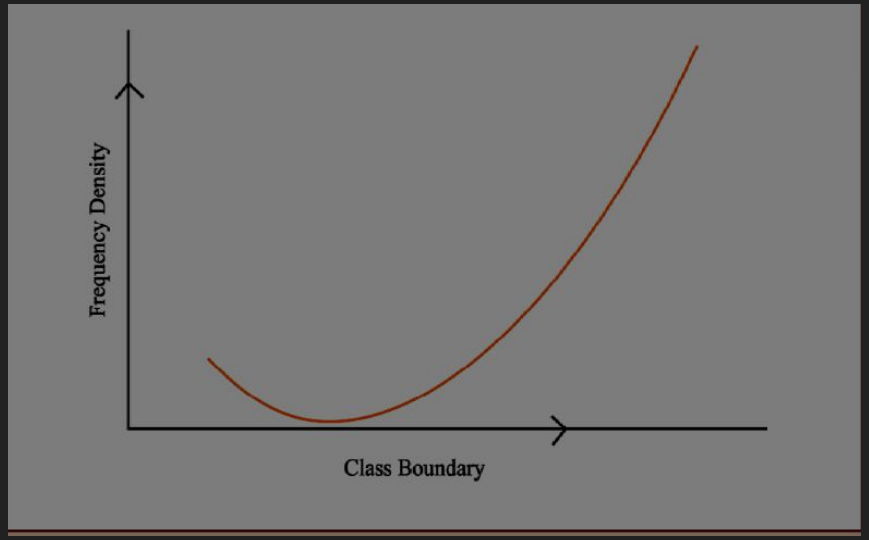
U - SHAPED CURVE

For a U-shaped curve , the frequency is minimum near the central part and the frequency slowly but steadily reaches its maximum at the two extremities .

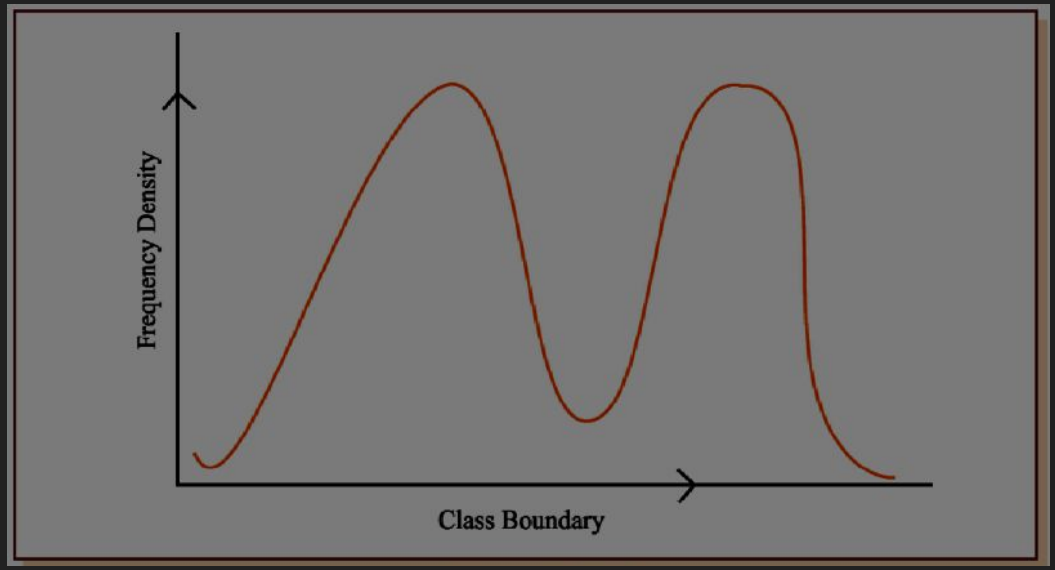


J - SHAPED CURVE

The J-shaped curve starts with a minimum frequency and then gradually reaches its maximum frequency at the other extremity .



MIXED CURVE



Que 52. For a particular class boundary, the less than cumulative frequency and more than cumulative frequency add up to

(a) Total frequency

(b) Fifty per cent of the total frequency

(c) (a) or (b)

(d) None of these.

a

Que 53. Frequency density corresponding to a class interval is the ratio of

(a) Class frequency to the total frequency

(b) Class frequency to the class length

(c) Class length to the class frequency

(d) Class frequency to the cumulative frequency.

b

Que 54. Relative frequency for a particular class

(a) Lies between 0 and 1

(b) Lies between 0 and 1, both inclusive

(c) Lies between –1 and 0

(d) Lies between –1 to 1.

a

Que 55. Mode of a distribution can be obtained from

(a) Histogram

(b) Less than type ogives

(c) More than type ogives

(d) Frequency polygon.

a

Que 56. Median of a distribution can be obtained from

(a) Frequency polygon

(b) Histogram

(c) Less than type ogives

(d) None of these.

C

Que 57. A comparison among the class frequencies is possible only in

(a) Frequency polygon

(b) Histogram

(c) Ogives

(d) (a) or (b).

b

Que 58. Frequency curve is a limiting form of

(a) Frequency polygon

(b) Histogram

(c) (a) or (b)

(d) (a) and (b).

C

Que 59. Most of the commonly used frequency curves are

(a) Mixed

(b) Inverted J-shaped

(c) U-shaped

(d) Bell-shaped.

d

Que 60. The distribution of profits of a company follows

(a) J-shaped frequency curve

(b) U-shaped frequency curve

(c) Bell-shaped frequency curve

(d) Any of these.

C

SET B

Que. 1 Out of 1000 persons, 25 per cent were industrial workers and the rest were agricultural workers. 300 persons enjoyed world cup matches on TV. 30 per cent of the people who had not watched world cup matches were industrial workers. What is the number of agricultural workers who had enjoyed world cup matches on TV?

(a) 260

(b) 240

(c) 230

(d) 250

Que. 2 A sample study of the people of an area revealed that total number of women were 40% and the percentage of coffee drinkers were 45 as a whole and the percentage of male coffee drinkers was 20. What was the percentage of female non-coffee drinkers?

(a) 10

(b) 15

(c) 18

(d) 20

Que. 3 Cost of sugar in a month under the heads raw materials, labour, direct production and others were 12, 20, 35 and 23 units respectively. What is the difference between the central angles for the largest and smallest components of the cost of sugar?

(a) 72°

(b) 48°

(c) 56°

(d) 92°

Que. 4 The number of accidents for seven days in a locality are given below :

No. of accidents :	0	1	2	3	4	5	6
Frequency :	15	19	22	31	9	3	2

What is the number of cases when 3 or less accidents occurred?

(a) 56

(b) 6

(c) 68

(d) 87

Que. 5 The following data relate to the incomes of 86 persons :

Income in Rs. :	500–999	1000–1499	1500–1999	2000–2499
No. of persons :	15	28	36	7

What is the percentage of persons earning more than Rs. 1500?

(a) 50

(b) 45

(c) 40

(d) 60

Que. 6 The following data relate to the marks of a group of students:

Marks :	Below 10	Below 20	Below 30	Below 40	Below 50
No. of students :	15	38	65	84	100

How many students got marks more than 30?

(a) 65

(b) 50

(c) 35

(d) 43

Que. 7 Find the number of observations between 250 and 300 from the following data :

Value :	More than 200	More than 250	More than 300	More than 350
No. of observations :	56	38	15	0

(a) 56

(b) 23

(c) 15

(d) 8

SAMPLING

POPULATION / UNIVERSE

- All items ,elements , or observations of interest having similar properties are known as population .
- It may be defined as the aggregate of all the units under consideration .

- **Example :** Population of students enrolled for CA Course

- The number of units belonging to a population is known as population size(N).

POPULATION

- If a population comprises only a finite number of units, then it is known as a **finite population.**
- **EXAMPLE :** Population of students enrolled for CA Course

- If the population contains an infinite or uncountable number of units, then it is known as an **infinite population.**
- **EXAMPLE :** population of stars, the population of mosquitoes

- A population consisting of real objects is known as an **existent population.**

- A population that exists just hypothetically like the population of heads when a coin is tossed infinitely is known as a hypothetical or an **imaginary population**

Census

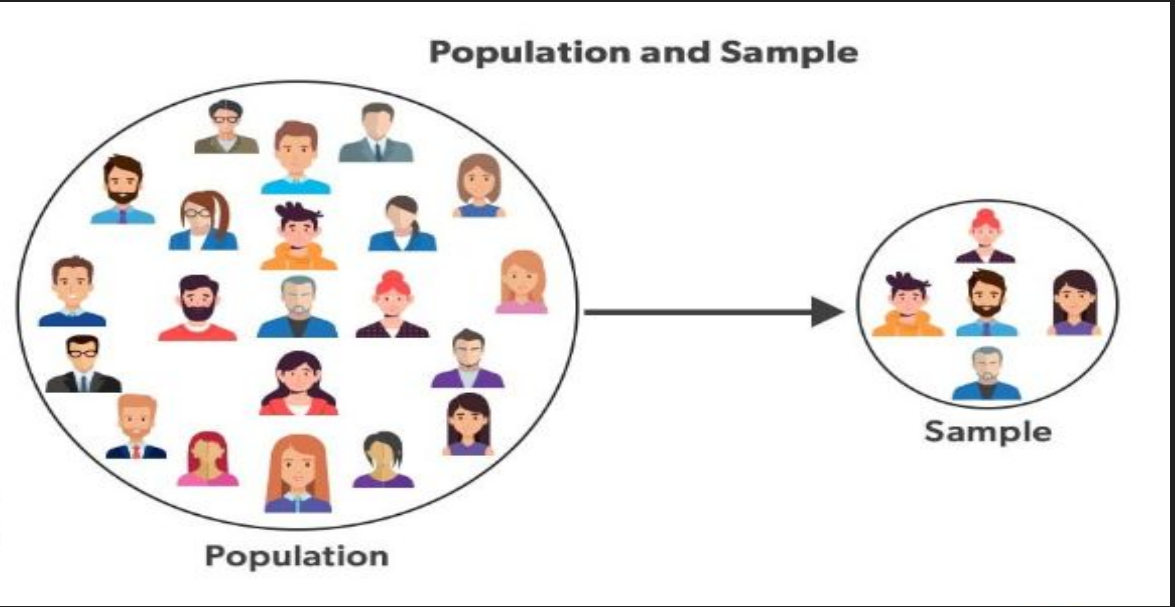
- Study of every elements of population is called census .

SAMPLE

- A sample may be defined as a part of a population so selected with a view to representing the population in all its characteristics .

- If a sample contains n units, then n is known as sample size.

- The units forming the sample are known as “Sampling Units”.
- A detailed and complete list of all the sampling units is known as a “Sampling Frame”.


Population and Sample
Population
Sample

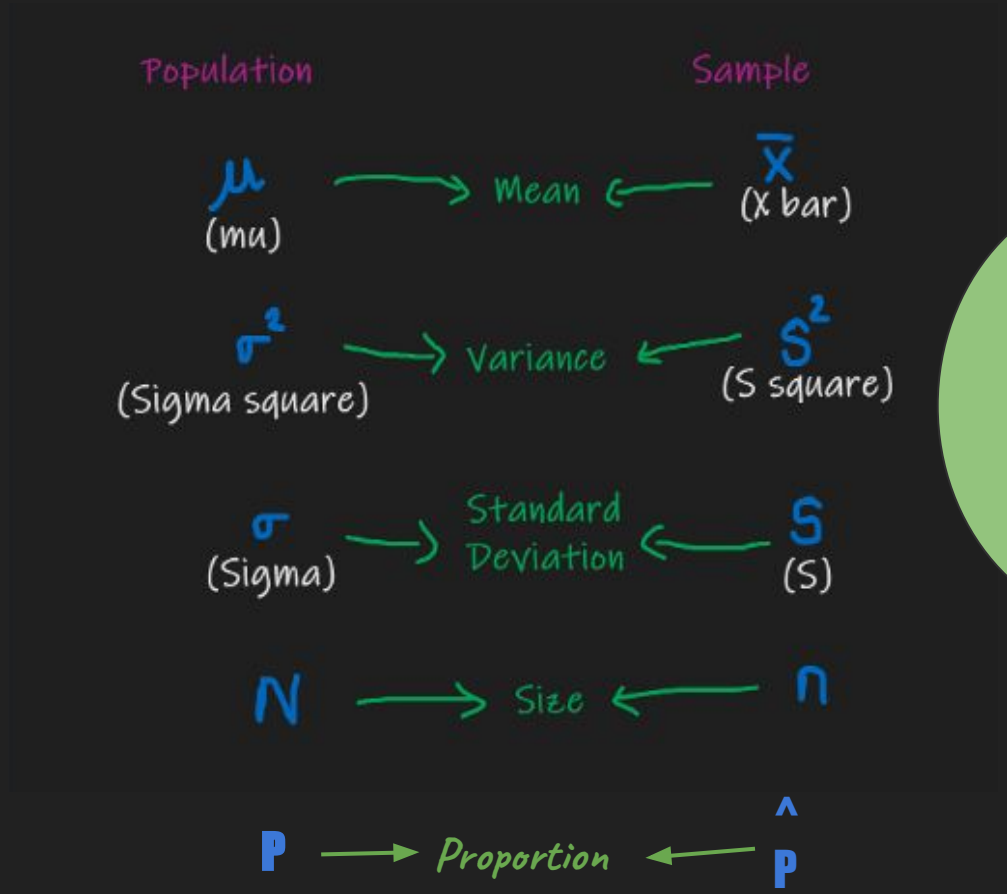
There are different statistical measures in statistics such as mean , median , mode , standard deviation , variance , proportion etc . These can be computed for both population and sample .

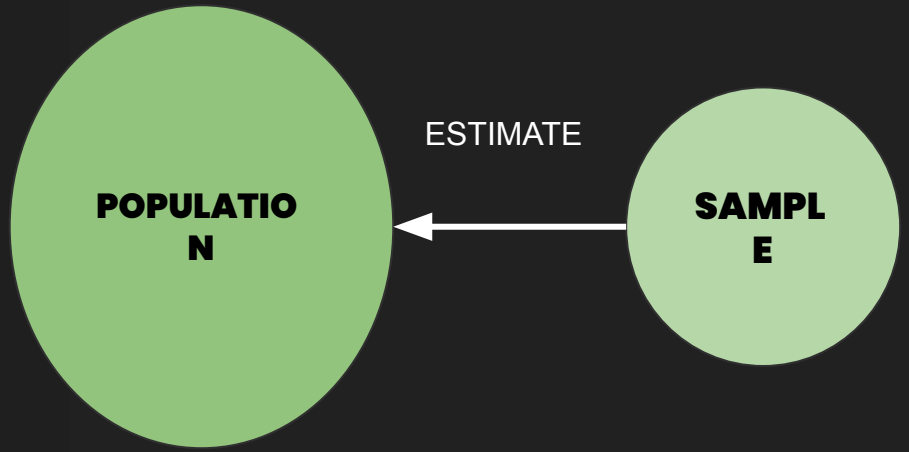
PARAMETER

- It is the statistical measures computed from population.
- A parameter may be defined as a characteristic of a population based on all the units of the population

STATISTIC

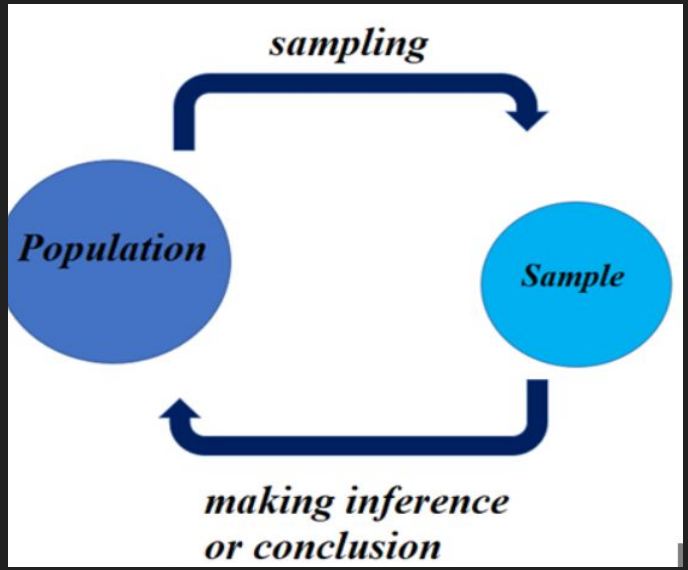
- It is the statistical measures computed from Sample .
- A statistic may be defined as a statistical measure of sample observation and as such it is a function of sample observations



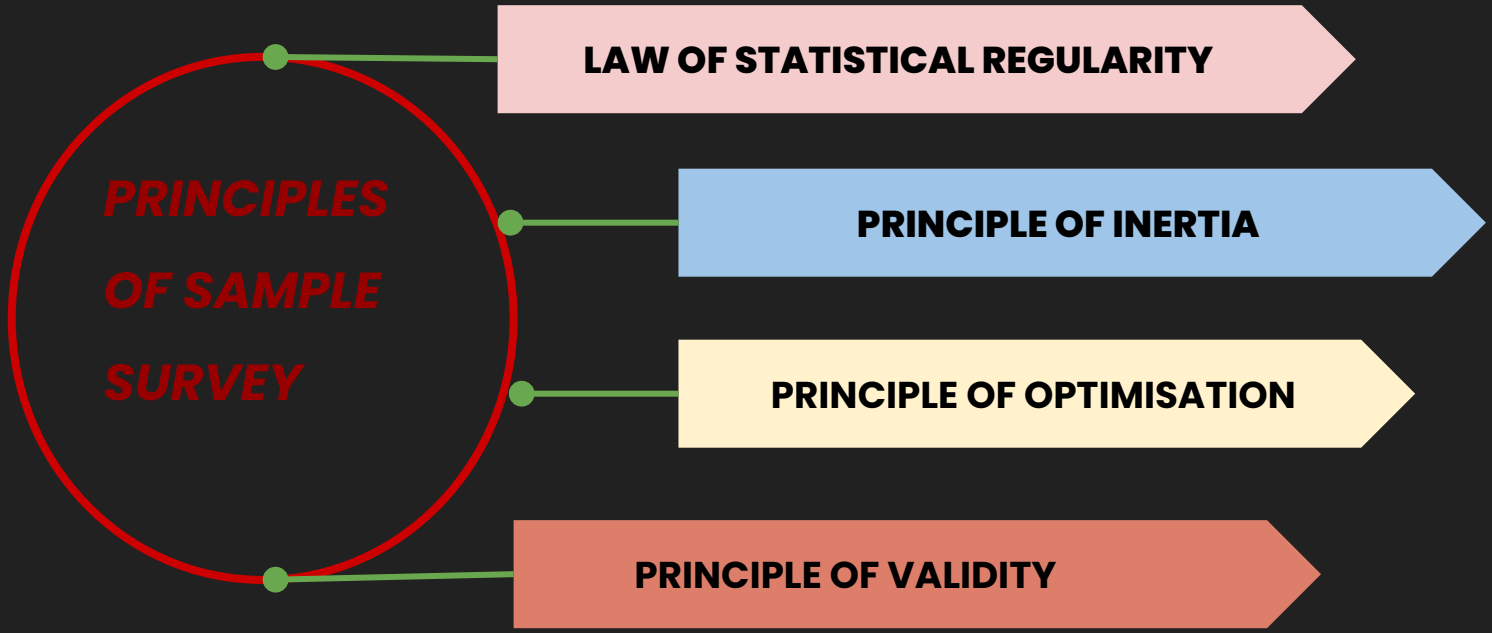


- *A statistic is used to estimate a particular population parameter*

- Sampling is a technique of selecting individual members or subset of the population to make statistical inferences from them and estimate characteristics of the whole universe .



- Sample Survey is the study of the unknown population on the basis of a proper representative sample drawn from it .



- **LAW OF STATISTICAL REGULARITY :** According to the law of statistical regularity, if a sample of fairly large size is drawn from the population under discussion at random, then on an average the sample would possess the characteristics of that population.

- **PRINCIPLE OF INERTIA :** It states that as sample size increases , the results are likely to be more reliable , accurate and precise , provided other factors are kept constant

- **PRINCIPLE OF OPTIMISATION :** The principle of optimization ensures that an optimum level of efficiency at a minimum cost or the maximum efficiency at a given level of cost can be achieved with the selection of an appropriate sampling design.

- **PRINCIPLE OF VALIDITY :** The principle of validity states that a sampling design is valid only if it is possible to obtain valid estimates and valid tests about population parameters.
- Only a probability sampling ensures this validity.

COMPARISON BETWEEN SAMPLE SURVEY AND COMPLETE ENUMERATION

- When complete information is collected for all the units belonging to a population, it is defined as complete enumeration or census.

- In most cases, we prefer sample survey to complete enumeration due to the following factors:

Speed: As compared to census, a sample survey could be conducted, usually, much more quickly simply because in sample survey, only a part of the vast population is enumerated.

COMPARISON BETWEEN SAMPLE SURVEY AND COMPLETE ENUMERATION

Cost : **The cost of collection of data on each unit in case of sample survey is likely to be more as compared to census because better trained personnel are employed for conducting a sample survey.**

But when it comes to total cost, sample survey is likely to be less expensive as only some selected units are considered in a sample survey.

COMPARISON BETWEEN SAMPLE SURVEY AND COMPLETE ENUMERATION

Reliability : The data collected in a sample survey are likely to be more reliable than that in a complete enumeration because of trained enumerators better supervision and application of modern technique.

COMPARISON BETWEEN SAMPLE SURVEY AND COMPLETE ENUMERATION

Accuracy : **Every sampling is subjected to what is known as sampling fluctuation which is termed as sampling error.**

It is obvious that complete enumeration is totally free from this sampling error.

It may be noted that in sample survey, the sampling error can be reduced to a great extent by taking several steps like increasing the sample size, adhering to a probability sampling design strictly and so on.

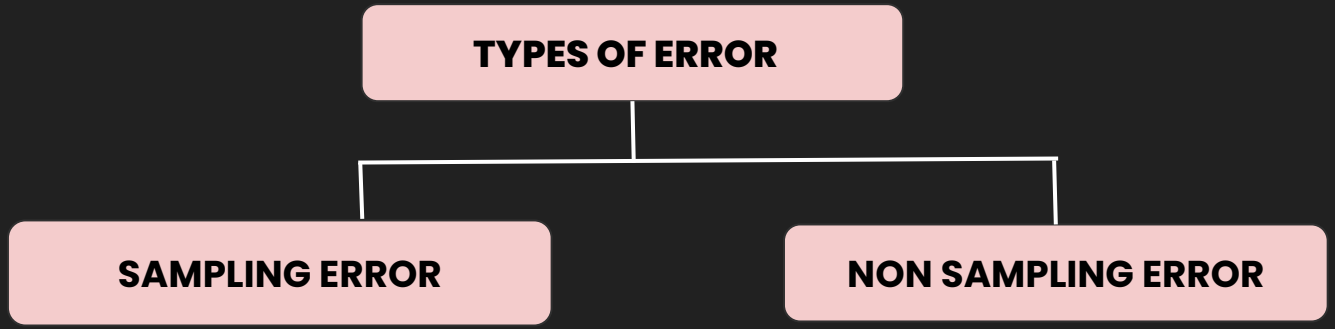
COMPARISON BETWEEN SAMPLE SURVEY AND COMPLETE ENUMERATION

Necessity : **Sometimes, sampling becomes necessity. When it comes to destructive sampling where the items get exhausted like testing the length of life of electrical bulbs or sampling from a hypothetical population like coin tossing, there is no alternative to sample survey.**

However, when it is necessary to get detailed information about each and every item constituting the population, we go for complete enumeration.

ERRORS IN SAMPLE SURVEY

- Errors or biases in a survey may be defined as the deviation between the value of population parameter as obtained from a sample and its observed value.



SAMPLING ERROR

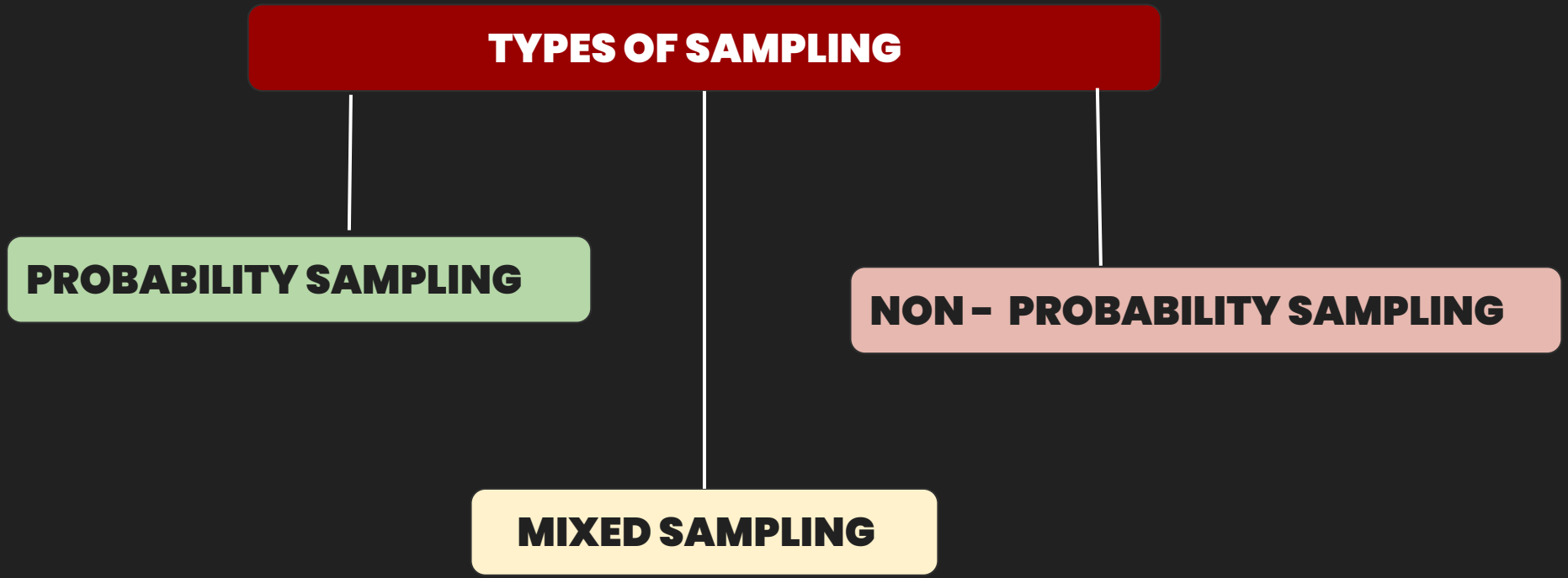
- Since only a part of population is investigated in sampling , every sampling design is subjected to this type of errors . Sampling errors are absent in census survey .
- **Factors contributing to sampling errors are as follows :**
- Errors arising out due to defective sampling design:
- *Errors arising out due to substitution*
- Errors owing to faulty demarcation of units:
- Errors owing to wrong choice of statistics :
- Variability in the population:

NON SAMPLING ERROR

- Errors due to recording observations, biases on the part of the enumerators, wrong and faulty interpretation of data is termed as non-sampling errors.
- This type of errors happen both in sampling and complete enumeration

Factors contributing to Non sampling errors are as follows :

- Lapse of memory
- Ignorance
- Communication gap
- Faulty planning
- Errors in compilation
- Non response bias
- Incomplete coverage


TYPES OF SAMPLING
PROBABILITY SAMPLING
MIXED SAMPLING
NON - PROBABILITY SAMPLING

PROBABILITY SAMPLING

- In the Probability sampling there is always a fixed, pre assigned probability for each member of the population to be a part of the sample taken from that population

- **Some important probability sampling are :**

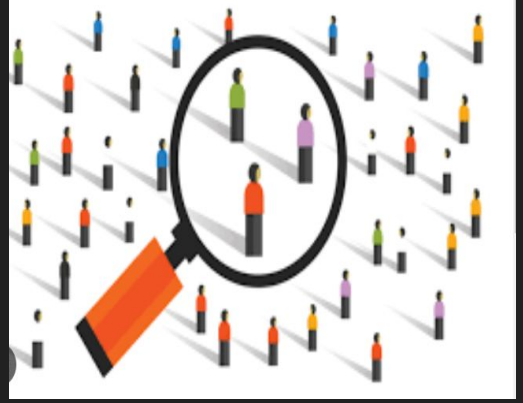
  simple random sampling ,

  stratified sampling,

  Multi Stage sampling, Multi Phase Sampling, Cluster Sampling and so on.

SIMPLE RANDOM SAMPLING

- When the units are selected independent of each other in such a way that each unit belonging to the population has an equal chance of being a part of the sample, the sampling is known as Simple random sampling or just random sampling.



SIMPLE RANDOM SAMPLING IS EFFECTIVE IF

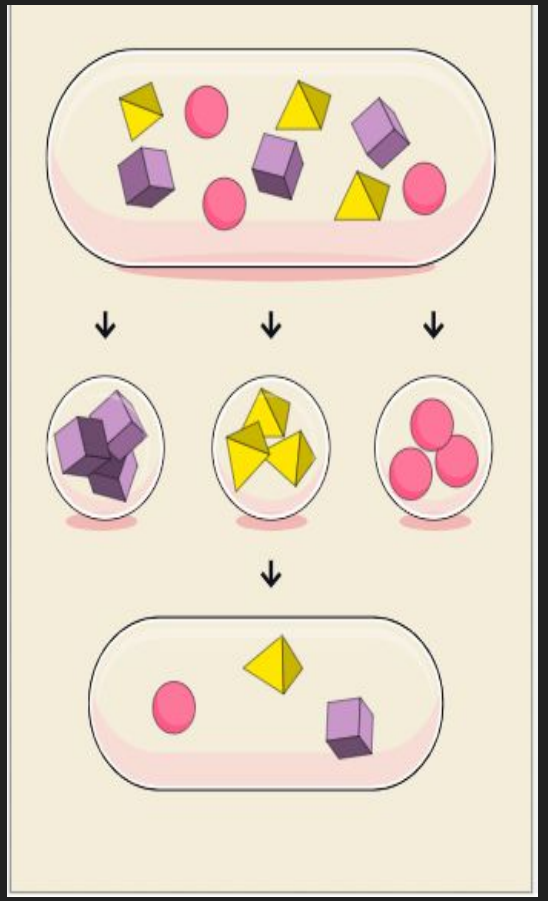
- the population is not very large
- the sample size is not very small
- the population under consideration is not heterogeneous

STRATIFIED SAMPLING

- In this method , the universe or the entire population is divided into a number of groups or strata and then certain number of items are taken from each group at random .

- Its basic purpose is to ensure that all the characteristics of a heterogeneous population are adequately represented in the sample .

- It helps in reduction of variability and thereby an increase in precision.



- There are two types of allocation of sample size.

- "Proportional allocation" or "Bowely's allocation
- When there is not much variation between the strata variances
- sample sizes for different strata are taken as proportional to the population sizes.

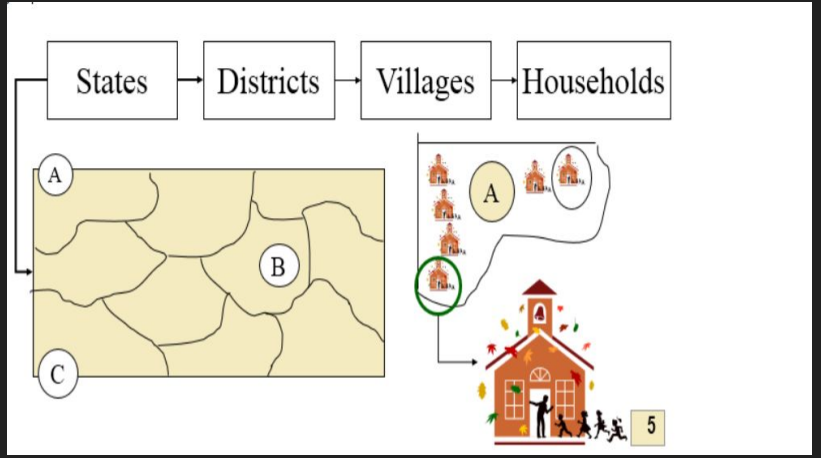
- "Neyman's allocation"
- When the strata-variances differ significantly among themselves
- sample size vary jointly with population size and population standard deviation

STRATIFIED SAMPLING

- **The purpose of stratified sampling are**
- (i) to make representation of all the sub populations
- (ii) to provide an estimate of parameter not only for all the strata but also and overall estimate
- (iii) reduction of variability and thereby an increase in precision.
- **Stratified sampling not advisable if**
- (i) population is not large
- (ii) some prior information is not available
- (iii) there is not much heterogeneity among the units of population

MULTISTAGE SAMPLING

- In this type of complicated sampling, the population is supposed to compose of first stage sampling units, each of which in its turn is supposed to compose of second stage sampling units, each of which again in its turn is supposed to compose of third stage sampling units and so on till we reach the ultimate sampling unit.



Suppose we want to take a sample of 5000 households from India

MULTISTAGE SAMPLING

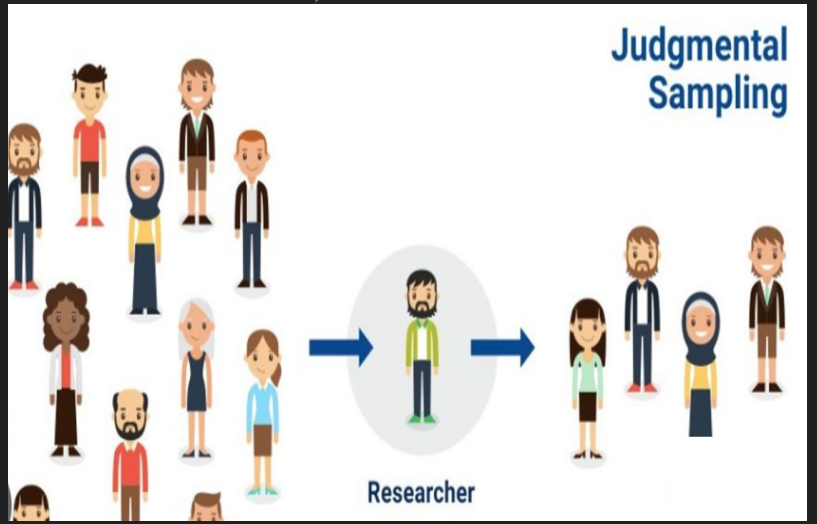
- The coverage in case of multistage sampling is quite large.
- It also saves computational labour and is cost-effective.
- It adds flexibility into the sampling process which is lacking in other sampling schemes.
- However, compared to stratified sampling, multistage sampling is likely to be less accurate.

NON - PROBABILITY SAMPLING

- In non- probability sampling , no probability attached to the member of the population and as such it is based entirely on the judgement of the sampler.
- Non-probability sampling is also known as Purposive or Judgemental Sampling

PURPOSIVE OR JUDGEMENTAL SAMPLING

- This type of sampling is dependent solely on the discretion of the sampler and he applies his own judgement based on his belief, prejudice, whims and interest to select the sample.
- Since this type of sampling is non-probabilistic, it is purely subjective and, as such, varies from person to person.
- No statistical hypothesis can be tested on the basis of a purposive sampling

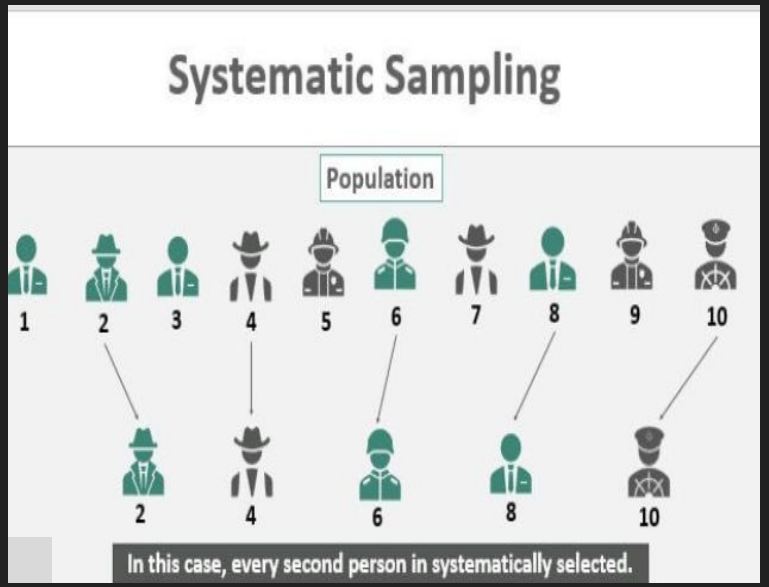


MIXED SAMPLING

- Mixed sampling is based partly on some probabilistic law and partly on some pre decided rule.
- Systematic sampling belongs to this category.

SYSTEMATIC SAMPLING

- It refers to a sampling scheme where the units constituting the sample are selected at regular interval after selecting the very first unit at random i.e., with equal probability.
- Systematic sampling is partly probability sampling in the sense that the first unit of the systematic sample is selected probabilistically and partly non- probability sampling in the sense that the remaining units of the sample are selected according to a fixed rule which is non-probabilistic in nature.



SYSTEMATIC SAMPLING

- If the population size N is a multiple of the sample size n i.e. **N = nk**, for a positive integer k which must be less than n, then the systematic sampling comprises selecting one of the first k units at random, usually by using random sampling number and thereby selecting every k^{th} unit till the complete, adequate and updated sampling frame comprising all the members of the population is exhausted. This type of systematic sampling is known as "***linear systematic sampling***". K is known as "***sample interval***".

SYSTEMATIC SAMPLING

- However, if N is not a multiple of n, then we may write $N = nk + p$, $p < k$ and as before, we select the first unit from 1 to k by using random sampling number and thereafter selecting every kth unit in a cyclic order till we get the sample of the required size n. This type of systematic sampling is known as "***circular systematic sampling***."

SAMPLING FLUCTUATION

It is the variation in the value of a statistic computed from different samples .

- **If we compute the value of a statistic, say mean, it is quite natural that the value of the sample mean may vary from sample to sample as the sampling units of one sample may be different from that of another sample.**

SAMPLING DISTRIBUTION

It is the probability distribution of a given statistic

- The mean of the statistic, as obtained from its sampling distribution, is known as “Expectation” and the standard deviation of the statistic is known as the “Standard Error (SE)“ .

SAMPLING DISTRIBUTION AND STANDARD ERROR OF STATISTIC

- SE can be regarded as a measure of precision achieved by sampling.
- SE is inversely proportional to the square root of sample size.

SAMPLING DISTRIBUTION AND STANDARD ERROR OF STATISTIC

- Starting with a population of N units, we can draw many a sample of a fixed size n.
- In case of **sampling with replacement**, the **total number of samples** that can be drawn is N^n
- When it comes to **sampling without replacement** , the **total number of samples** that can be drawn is ${}^{N}C_{n}$

Answer the following questions. Each question carries one mark.

Que. 1 Sampling can be described as a statistical procedure

(a) To infer about the unknown universe from a knowledge of any sample

(b) To infer about the known universe from a knowledge of a sample drawn from it

(c) To infer about the unknown universe from a knowledge of a random sample drawn from it

(d) Both (a) and (b).

C

Answer the following questions. Each question carries one mark.

Que. 2 The Law of Statistical Regularity says that

(a) Sample drawn from the population under discussion possesses the characteristics of the population

(b) A large sample drawn at random from the population would possess the characteristics of the population

(c) A large sample drawn at random from the population would possess the characteristics of the population on an average

(d) An optimum level of efficiency can be attained at a minimum cost.

C

Answer the following questions. Each question carries one mark.

Que. 3 A sample survey is prone to

(a) Sampling errors

(b) Non-sampling errors

(c) Either (a) or (b)

(d) Both (a) and (b)

d

Answer the following questions. Each question carries one mark.

Que. 4 The population of roses in Salt Lake City is an example of

(a) A FInite population

(b) An infinite population

(c) A hypothetical population

(d) An imaginary population.

b

Answer the following questions. Each question carries one mark.

Que. 5 **Statistical decision about an unknown universe is taken on the basis of**

(a) Sample observations

(b) A sampling frame

(c) Sample survey

(d) Complete enumeration

a

Answer the following questions. Each question carries one mark.

Que. 6 **Random sampling implies**

(a) Haphazard sampling

(b) Probability sampling

(c) Systematic sampling

(d) Sampling with the same probability for each unit.

d

Answer the following questions. Each question carries one mark.

Que. 7 **A parameter is a characteristic of**

(a) Population

(b) Sample

(c) Both (a) and (b)

(d) (a) or (b)

a

Answer the following questions. Each question carries one mark.

Que. 8 A statistic is

(a) A function of sample observations

(b) A function of population units

(c) A characteristic of a population

(d) A part of a population.

a

Answer the following questions. Each question carries one mark.

Que. 9 **Sampling Fluctuations may be described as**

(a) The variation in the values of a statistic

(b) The variation in the values of a sample

(c) The differences in the values of a parameter

(d) The variation in the values of observations.

a

Answer the following questions. Each question carries one mark.

Que. 10 **The sampling distribution is**

(a) The distribution of sample observations

(b) The distribution of random samples

(c) The distribution of a parameter

(d) The probability distribution of a statistic.

d

Answer the following questions. Each question carries one mark.

Que. 11 Standard error can be described as

(a) The error committed in sampling

(b) The error committed in sample survey

(c) The error committed in estimating a parameter

(d) Standard deviation of a statistic.

d

Answer the following questions. Each question carries one mark.

Que. 12 A measure of precision obtained by sampling is given by

(a) Standard error

(b) Sampling fluctuation

(c) Sampling distribution

(d) Expectation.

a

Answer the following questions. Each question carries one mark.

Que. 13 As the sample size increases, standard error

(a) Increases

(b) Decreases

(c) Remains constant

(d) Decreases proportionally.

b

***Answer** the following questions. Each question carries one mark.*

Que. 14 **If from a population with 25 members, a random sample without replacement of 2 members is taken, the number of all such samples is**

(a) 300

(b) 625

(c) 50

(d) 600

a

Answer the following questions. Each question carries one mark.

Que. 15 **A population comprises 5 members. The number of all possible samples of size 2 that can be drawn from it with replacement is**

(a) 100

(b) 15

(c) 125

(d) 25

d

Answer the following questions. Each question carries one mark.

Que. 16 **Simple random sampling is very effective if**

(a) The population is not very large

(b) The population is not much heterogeneous

(c) The population is partitioned into several sections.

(d) Both (a) and (b)

d

Answer the following questions. Each question carries one mark.

Que. 17 Simple random sampling is

(a) A probabilistic sampling

(b) A non- probabilistic sampling

(c) A mixed sampling

(d) Both (b) and (c).

a

Answer the following questions. Each question carries one mark.

Que. 18 According to Neyman's allocation, in stratified sampling

(a) Sample size is proportional to the population size

(b) Sample size is proportional to the sample SD

(c) Sample size is proportional to the sample variance

(d) Population size is proportional to the sample variance.

a

Answer the following questions. Each question carries one mark.

Que. 19 Which sampling provides separate estimates for population means for different segments and also an over all estimate?

(a) Multistage sampling

(b) Stratified sampling

(c) Simple random sampling

(d) Systematic sampling

b

Answer the following questions. Each question carries one mark.

Que. 20 Which sampling adds flexibility to the sampling process?

(a) Simple random sampling

(b) Multistage sampling

(c) Stratified sampling

(d) Systematic sampling

b

Answer the following questions. Each question carries one mark.

Que. 21 Which sampling is affected most if the sampling frame contains an undetected periodicity?

(a) Simple random sampling

(b) Stratified sampling

(c) Multistage sampling

(d) Systematic sampling

d

Answer the following questions. Each question carries one mark.

Que. 22 Which sampling is subjected to the discretion of the sampler?

(a) Systematic sampling

(b) Simple random sampling

(c) Purposive sampling

(d) Quota sampling.

C

Answer the following questions. Each question carries one mark.

Que. 23 If a random sample of size 2 with replacement is taken from the population containing the units 3,6 and 1, then the samples would be

(a) (3, 6),(3, 1),(6, 1)

(b) (3, 3),(6, 6),(1, 1)

(c) (3, 3),(3, 6),(3, 1),(6, 6),(6, 3),(6, 1),(1, 1),(1, 3),(1, 6)

(d) (1, 1),(1, 3),(1, 6),(6, 1),(6, 2),(6, 3),(6, 6),(1, 6),(1, 1)

C

Answer the following questions. Each question carries one mark.

Que. 24 If a random sample of size two is taken without replacement from a population containing the units a,b,c and d then the possible samples are

(a) (a, b),(a, c),(a, d)

(b) (a, b),(b, c), (c, d)

(c) (a, b), (b, a), (a, c),(c,a), (a, d), (d, a)

(d) (a, b), (a, c), (a, d), (b, c), (b, d), (c,d)

d