CHAPTER- Statistical Description of Data

1. Introduction to Statistics

There are many views about the origin of statistics which are as follows:-

1. Latin Word → 'Status'
2. Italian Word → 'Statista'
3. German word → 'Statistik'
4. French word → 'Statistique'

(•) We are also thankful to 'Kautalya' who kept record of birth and death as well as some other precious record in his famous book 'Arthashastra' during Chandragupta reign. (In 4th century B.C)

(•) During the reign of Akbar. (16th century AD) We find statistical record on Agriculture in **Ain-i-Akbari** written by **Abu Fazal**

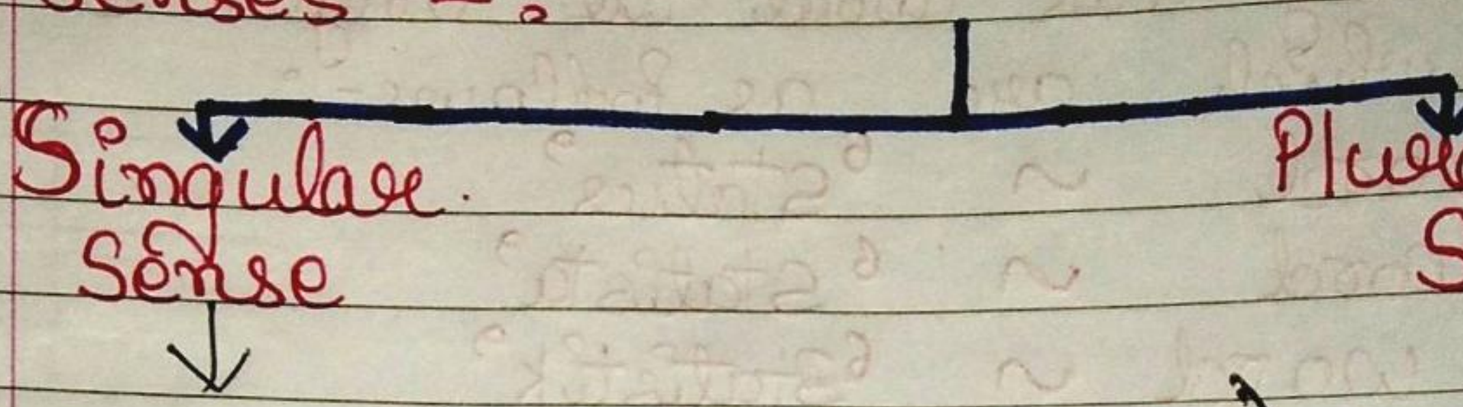
(•) Referring to egypt, the first census was conducted by the **Pharoh**. (During 3000 BC to 2000 B.C)

IMPORTANT NOTES TO KNOW ABOUT THE HISTORY OF STATISTICS

1. **P.C Mahalanobis** is an Indian statistician who made significant contribution in the development of statistics
2. The real giant in the development of statistics theory is **R.A. Fisher** ...

Defination of Statistics

We may define statistics ~~in a singular~~ in two senses :-



- It consist of all the scientific method that is employed for collecting analyzing, presenting data.
- It means 'science of counting' or 'science of averages'

- Croxton & Cowden has defined statistics in singular sense.

- When used plural it is defined as qualitative & quantitative that are collected usually for having analysis
- Prof Horace Secrist has defined statistics in plural sense.

Application of Statistics

Statistics has confined its applications into three basic feilds :-

1. Economics
   Time Series Analysis
   Index No.
   Demand Supply Analysis etc are some overlapping areas of Economics & Statistics.
- Econometrics → is the branch of stats with economics.

Page
Date / /

(2) Business Managment

Stats is an indespensable tool in all aspects of business. When a man enters business he enters business as he knows it is always forecasting which help business to run, this forecast is done through past trends which is judge by statistics. managed

(3) In COMMERCE & INDUSTRY.

1) Help in organization of business
2) Production.
3) Scientific Managment & Business forecasting
4) Purchase.

limitations of Statistics.

Statistics and its technique are widely used in every branch of knowledge.

W.I. King rightly says: "Science of statistics is the most useful servant. But only of great value to those who understand its proper use"

Important limitations

- INDIVIDUAL ITEM IGNORED
- Only deal with Quantitative data.
- Based on only averages
- Does not reveal entire story.
- Misused.
- Uniform & Homogenous

Collection of Data

- Data is some particulaer characteristic undeer consideration.

Quantitative Data.

- It is known as variable or in other words variable is a measureable quantity.

Variable

Discrete

When a variable assumes finite or countable infinite no. of isolated values it is Discrete Variable.

Eg.) No of Petals in a flower
② No of misprint in a books
③ No. of road accident in a particular locality and so. on.

Continous

If it can assume any value foroma given Interval. it is known as continous variable

Qualitative Data.

- A qualitative characteristics is known as attributes.

Eg. Gender, National. Colour, Caste.

We can brodly classify data as -:

a) Primary Data.
It is collection of original data for the first time.

Sources -:
1) Interview Method.
2) Mailed Questionnaire Method
3) Observation Method.
4) Questionere sent by Enumerator.

b) Secondry Data.
Basically it is compilation of existing data.

Sources -:
1) International source.
2) Govt source.
3) Published sour.

IMPORTANT POINTS

- In case of Natural calamity like a Supere cyclone. or an earthquake or an epidemic. like plague. we may use Interview. Method.
- In case of rail accident we may recourse for conducting Indirect Interviews.
- Telephone Interview method is a quick and rather non-expensive way to collect primary data
- Wide Area is covered using mailed questionere method, the amount of non-responses is maximum in this metho
- Data Collected on religion is from Census report is secondary data.
- Questionaire Method is used when informants are literate.
- When the field enquiry is wide then Questionaire sent through Enumerators of Best Method.

Note To Remember

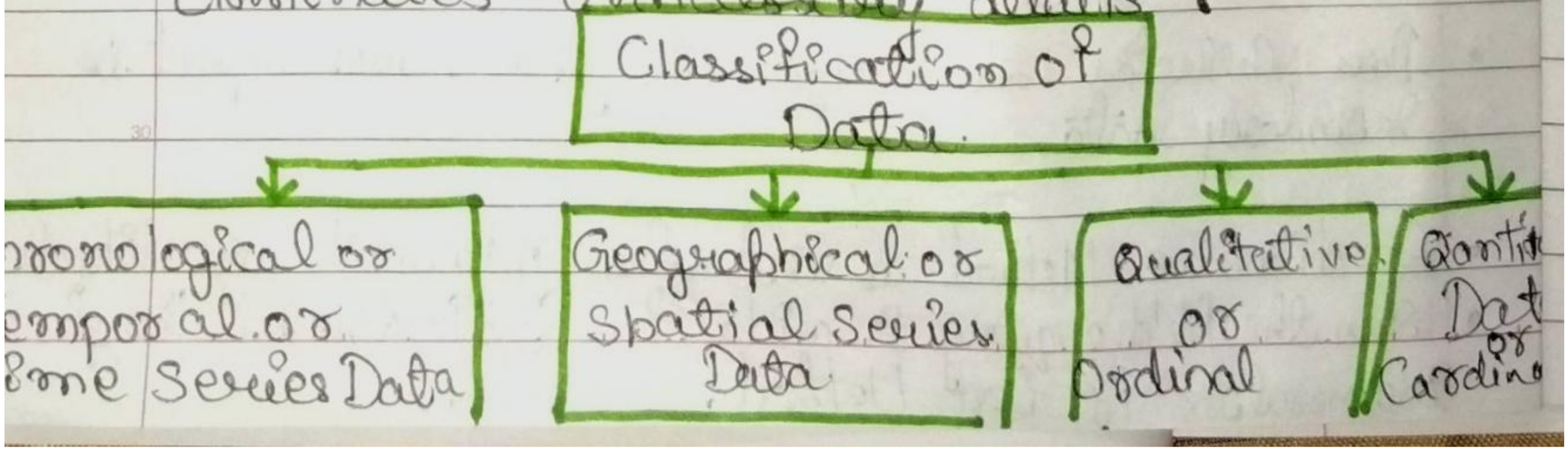
- The term error in Statics refer to bias.
- Sampling error are present only in sampling survey. and usually decreased with increase in sample size.
- Non-Sampling error include both bias and mistakes and may occur in Sample Survey and Enumeration Survey

Classification or Organisation of Data

Process of arranging data on the basis of characteristic into groups or classes

Why to classify data?

- To put data into neat, precise and condense form to be easily understood.
- To make comparision easy.
- Statistical analysis is possible only for the classified data.
- Eliminates unnecessary details.



• Data can be classified as frequency & Non-frequency data. The Qualitative & Quantitative only belong to frequency data.

Mode of presenting data.

1. Textual Presentation

Method of presenting data in the form of long paragraphs.

Merits:
- Simplicity.
- Exact Magnitude can be presented.
- It is first step to other method.

Demerits
- (•) Dull.
- (•) Monotonous
- (•) Comparision not possible

2. Tabular presentation

Presentation of data with the help of statistical table.

- A table has alloted Serial No. with self explanatory title.
- Caption is the uppermost part of the table describing columns & sub-columns
- Box-head is the entire upper part of the table which includes columns & sub-columns number, unit(s) of measurment.
- Stub is the left part of the table providing description of rows
- Body is the main part of the table containing numerical figure
- Notes describing the source of data & bringing clarity about any row or columns known as footnotes, should be shown at the bottom part of the table.

Important Notes

Camlin Page
Date . . ..

(1) A table showing height & weight is an example of quantitative classification.

(2) A table showing population of delhi is an example of Temporal classification.

(3) State wise wheat production is spatial classification.

(4) Mutually exclusive classification is meant for continous variable.

(5) Real valued function on infinite sample space is Continous random variable. Discrete random variable is a real valued function on finite sample.

(6) Exclusive class limit are suitable for continous variable & inclusive limit are suitable for discrete variable

(7) In exclusive series upper limit is not included, whereas upper limit is included in inclusive data.

(8) Distribution of share is an example of the frequency distribution of a discrete variable and the distribution of profit is an example of the frequency distribution of a continous variable.

(9) In tabulation row designation are called stubs and column heading captions

10. The unit of measurement in tabulation is shown in Box-head.

11. Middle Value = $\frac{\text{Upper limit} + \text{Lower limit}}{2}$.

(3) Digrammatic and Graphical Presentation

1.) One Dimensional diagrams.

These diagrams are prepared on the bases of length. This type of diagram take the shape of bars.

(a) SIMPLE BAR DIAGRAM

It is constructed to represent one value of a given variable, the length of various bars is in the ratio of the magnitude of the given data.

(b) MULTIPLE BAR DIAGRAM.

It is constructed to represent the value of different variable or the value of various component of the same variable.

(c) SUB-DIVIDED BAR DIAGRAM

It is diagram in which one bar is constructed for the total value of the variable and this bar is sub-divided in proportion of the values of various component to that variable.

2.) TWO-DIMENSIONAL DIAGRAMS (AREA DIAGRAM)

It is prepared on the basis of 2-D i.e length & breadth like the product of both is area, same is this area diagram

The following types are area diagram

1) Rectangle diagram
2) Sub-divided Rectangle diagram.
3) Square diagram
4) Circle diagram
5) Pie-diagrams

Types of Graphs

(1) Time-series graphs.

(a) ONE DEPENDENT Variable Histogram.
It is prepared to the value of one dependent variable over different period of time.

b) More than one dependent variable Histogram
It is prepared to show the value of more than one dependent variable over different to time.

(c) Mixed graph.
It is prepared to show the two dependent variable with two different units of measurement.

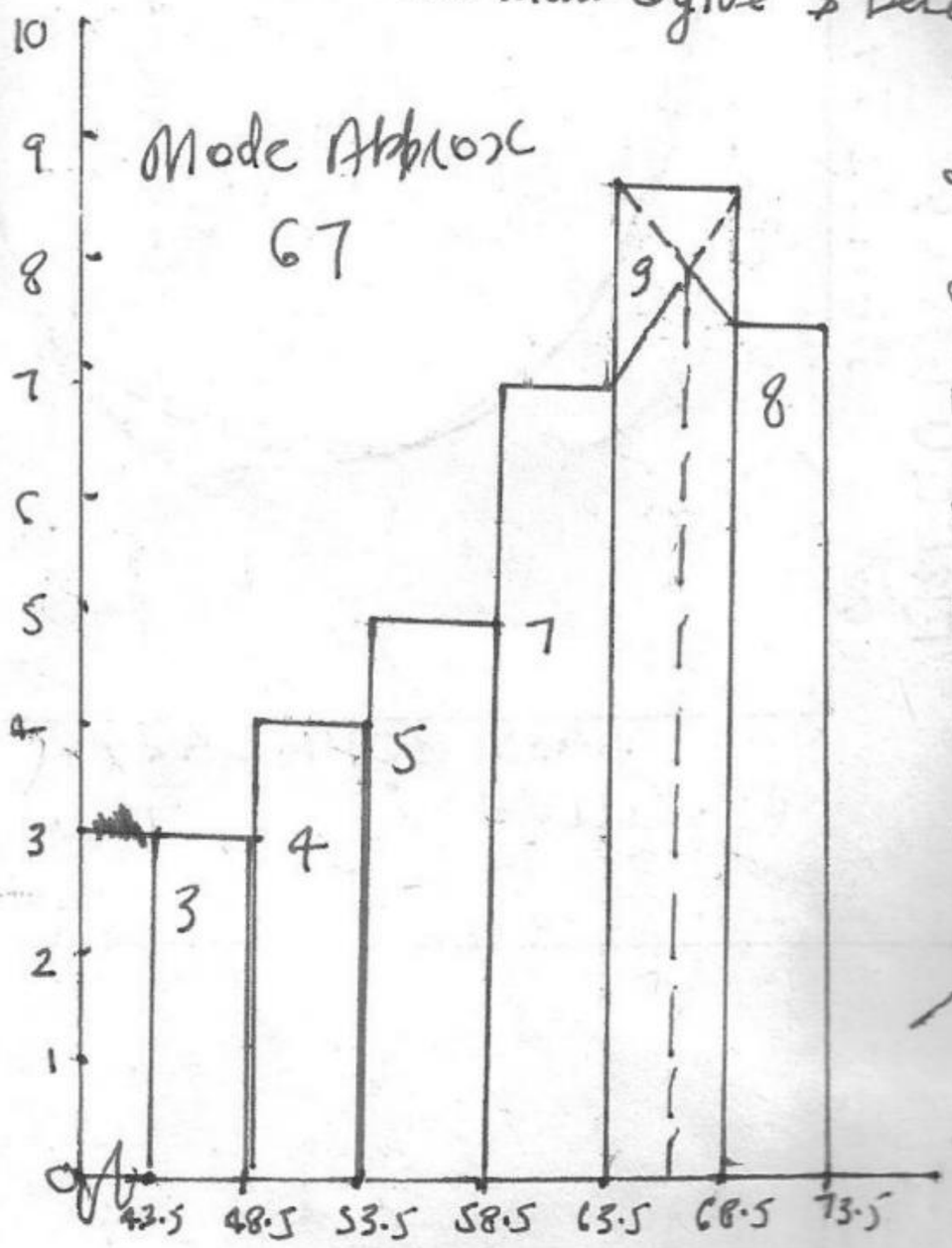
(d) Range Graph
It is a graph prepared to show the range of the data b/w two extremes value at different point of time.

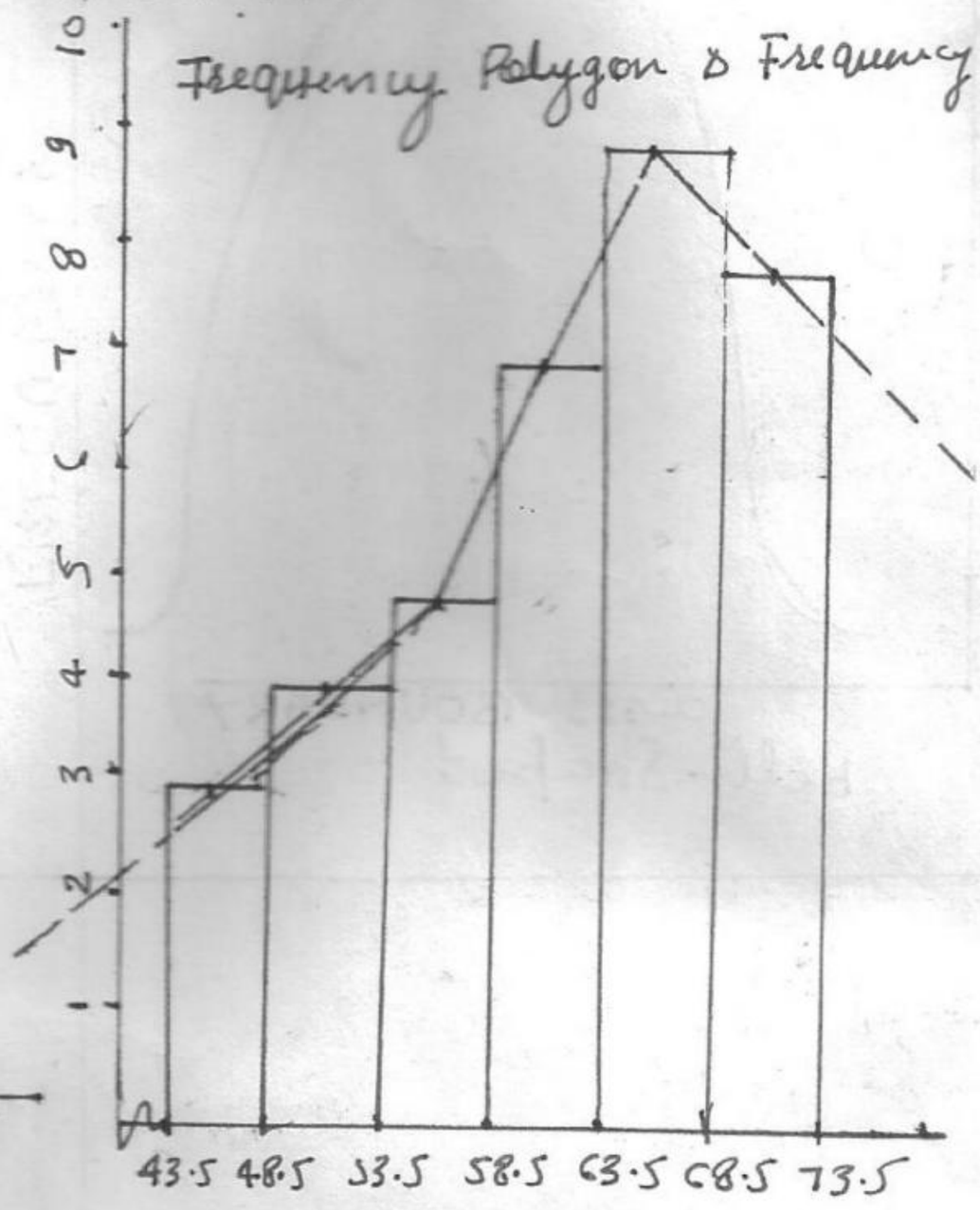
2. Frequency distribution Graphs.

a) Histogram.
b) Frequency polygon.
c) Frequency curve.
d) ogive.

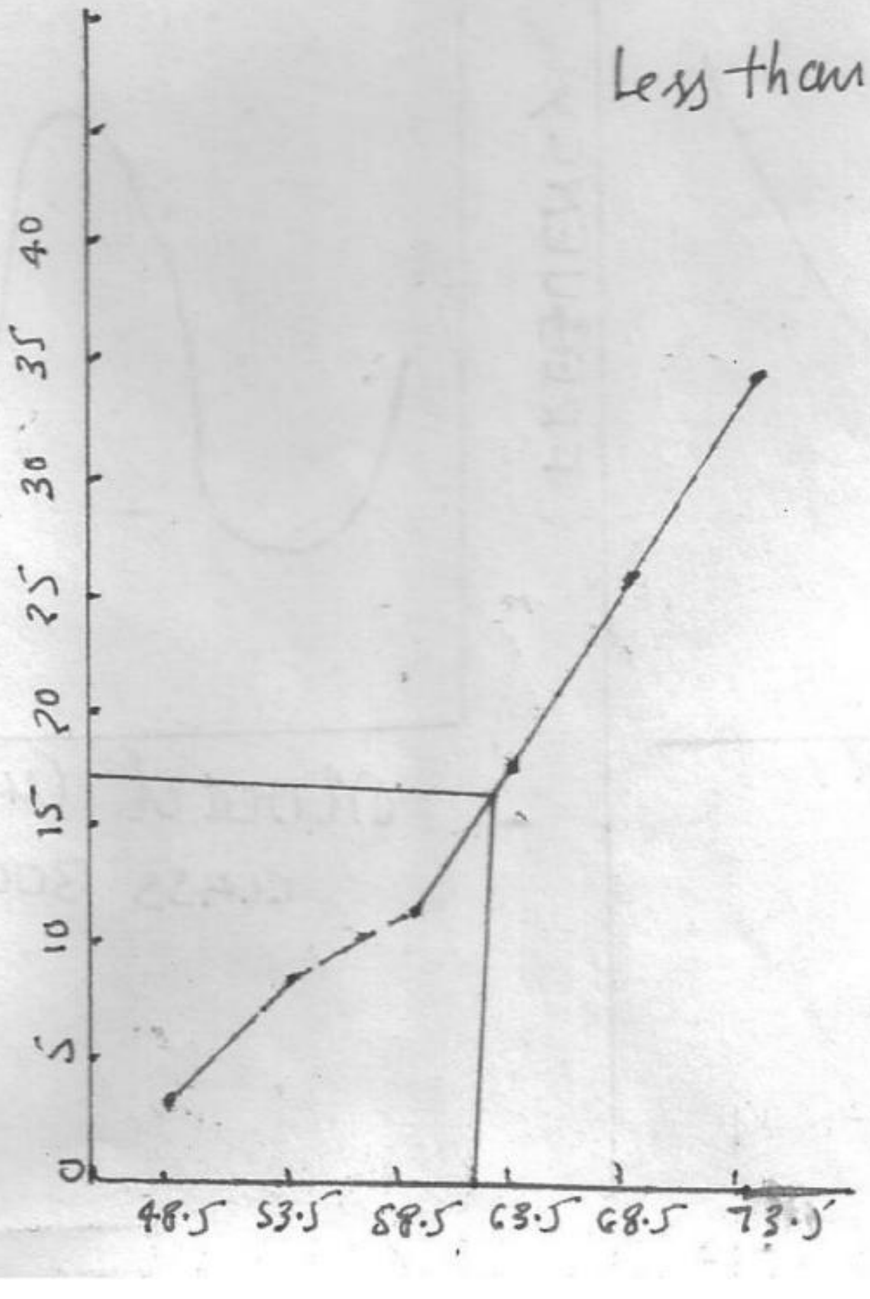
weight	No. of student	C.f
43.5 – 48.5	3	3
48.5 – 53.5	4	7
53.5 – 58.5	5	12
58.5 – 63.50	7	19
63.5 – 68.50	9	28
68.5 – 73.50	8	36

(i) Draw the Histogram & Find out mode.
(ii) Frequency polygon, Frequency Curve.
(iii) Draw less than ogive & Determine Median

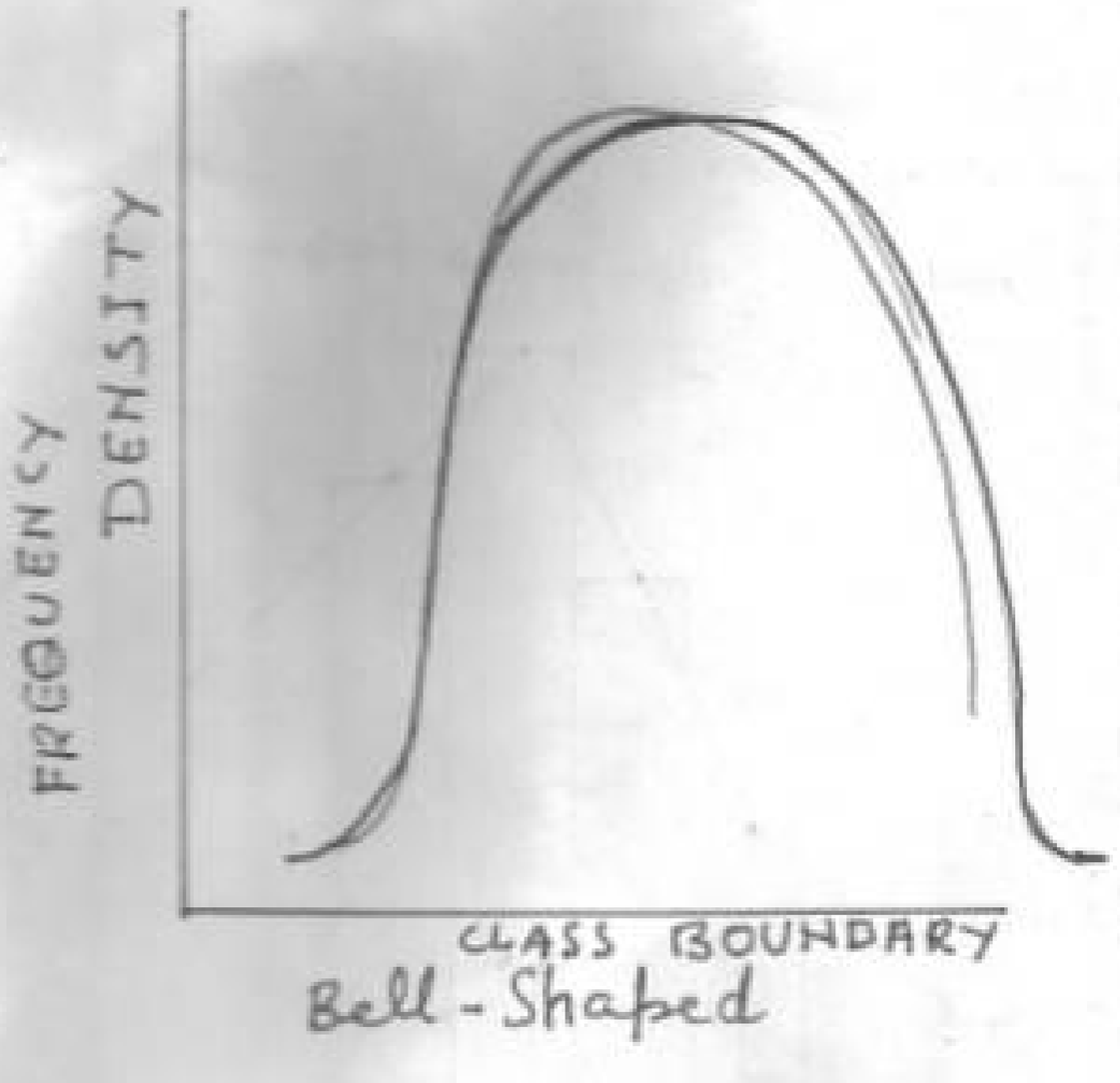




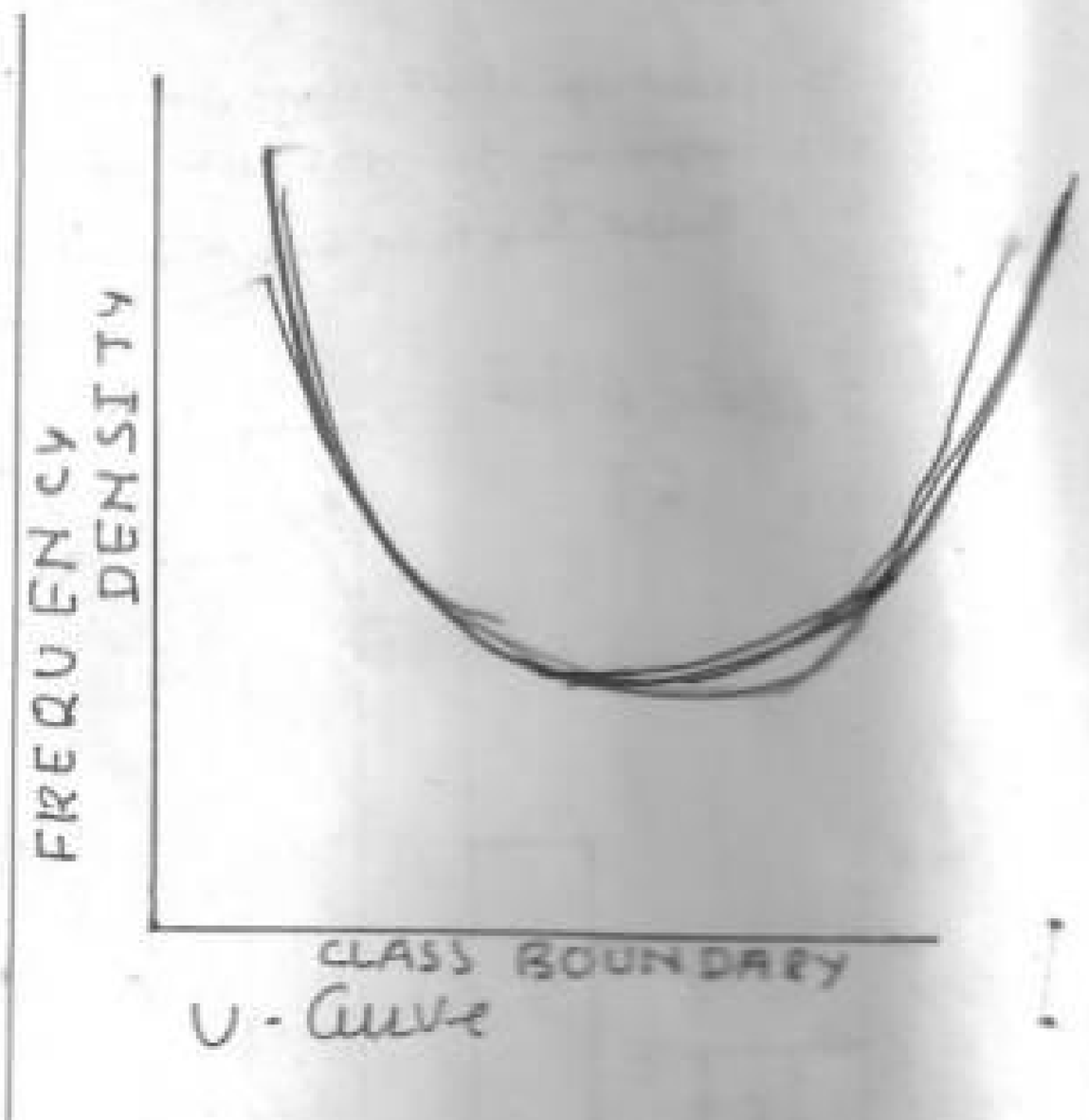
Less than ogive



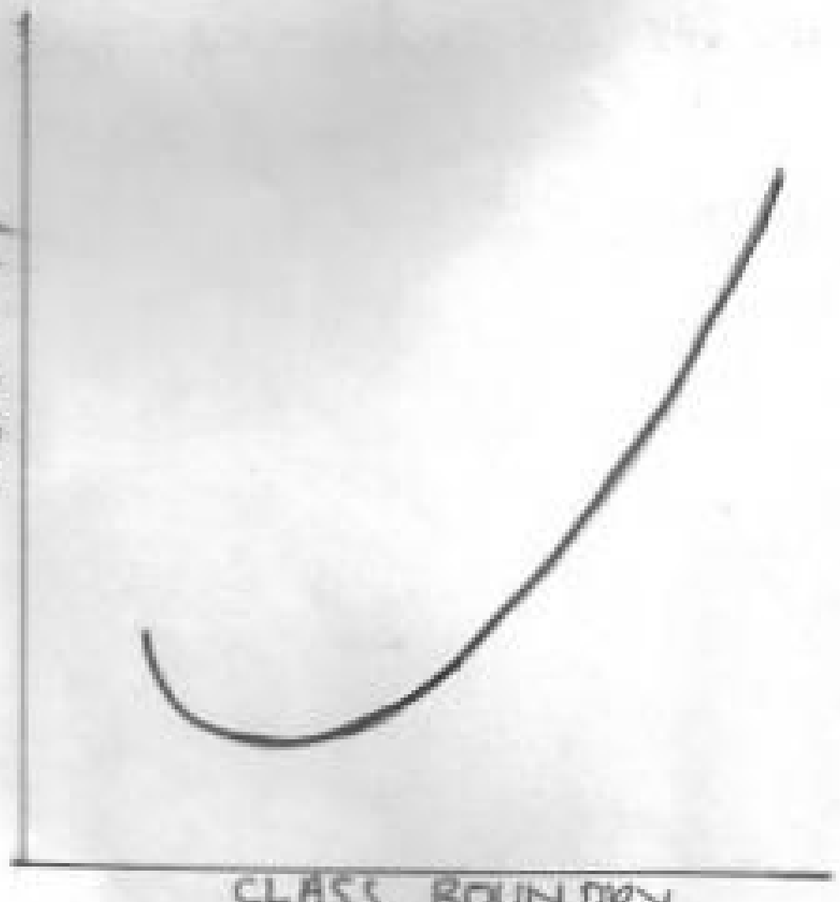
FREQUENCY CURVE



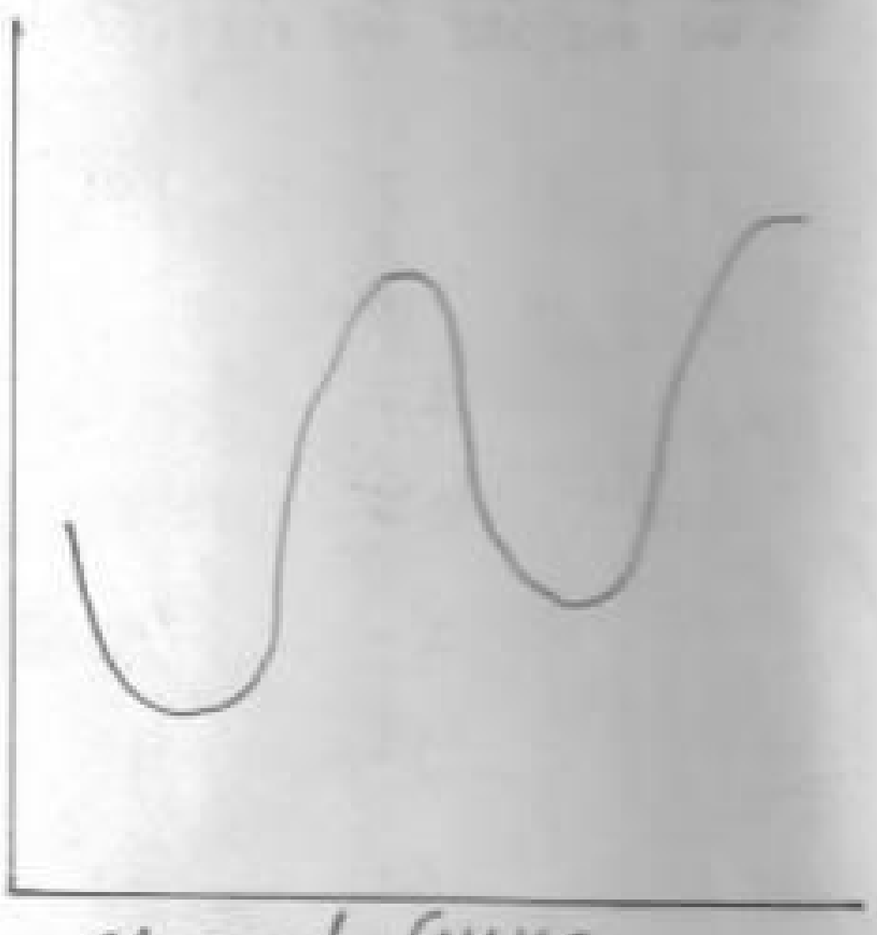
Bell-Shaped



U-Curve



J Shape



Mixed Curve