

GERMAN WORD → STATISTIK, FRENCH WORD - STATISTIQUE
 LATIN WORD → STATUS, ITALIAN WORD :- STATISTICA

Background of Chapter

Phase-1

Basics

- Introduction
- Importance of Stats
- Limitations of Stats

Data & Types

- Quantitative Data → Discrete, Continuous
- Qualitative Data → We assign the value

Method of Collection of Data

Classification of Data

Primary Source

- 1 Interview Method
- 2 Mailed Questionnaire Method
- 3 Observation Method
- 4 Questionnaire filled and sent by enumerators

Secondary Source

from all of them only International & Gov. are Important for Exam. Only these are Important.

- 1 International sources
- 2 Government sources
- 3 Private and quasi-government
- 4 Unpublished sources

Phase-2

Mode of Presentation of Data

- Textual presentation
- Tabular presentation
- Diagrammatic presentation

Frequency Distribution

Non-Graphical Methods

- Ascending / Descending
- Tabulation

Graphical Methods

- Histogram
- Frequency Polygon
- Frequency Curve
- Ogive

Line Diagram

- Single Line Chart
- Multiple Line Chart
- Multiple Line Axis Chart

Bar Diagram

- Simple Bar Diagram
- Multiple or Grouped Bar Diagram
- Divide or component Bar Diagram
- Percentage subdivided Bar Diagram

Pie Chart

Meaning and Definition of Statistics

Plural Sense

It refers to the **Numerical data (Qualitative as well as Quantitative)** collected in a systematic manner with some definite aim or object in view.

Example: The number of persons suffering from malaria in different colonies of Delhi or number of unemployed girls in different states of India and so on.

Singular Sense

Scientific method that is employed for collecting, analyzing and presenting data, leading finally to drawing statistical inferences about some important characteristics.

Application of statistics

1. Economics

Economics and Statistics are closely associated. Time Series Analysis, Index Numbers, Demand Analysis etc. are some overlapping areas of Economics and Statistics.

2. Business Management

Because of the never-ending complexity in the business and industry environment, most of the decision-making processes rely upon different quantitative techniques which could be described as a combination of statistical methods and operations research techniques.

3. Statistics in Commerce and Industry

Data on previous sales, raw materials, wages and salaries, products of identical nature of other factories etc. are collected, analyzed and experts are consulted in order to maximize profits. Measures of central tendency and dispersion, correlation and regression analysis, time series analysis, index numbers, sampling, statistical quality control are some of the statistical methods employed in commerce and industry.

Limitations of Statistics

Statistics and its techniques 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

1. Statistics does not deal with **individual item.**
2. Statistics deals with **quantitative data.**
3. Statistics laws are true **only on averages.**
4. Statistics does **not reveal the entire story.**
5. Statistics is liable to be **misused.**
6. Statistics data should be **uniform** and **homogeneous.**

Data & Types of Data

Data: The word data means **information** (its literary meaning is given as facts). This can be defined as the information collected through censuses and surveys or in a routine manner or other sources is called data.

Types of Data

Quantitative Data

Also known as a **variable** or in other words, a **variable is a measurable quantity.**

Qualitative Data

Also known as **attributes**. **Example:** Gender of a baby, Nationality of a person, Color of a flower etc.

Can not measure in numerical terms.

Discrete Variable: When a variable assumes a finite or a countably infinite number of isolated values.

- Examples:**
1. Number of misprints in a book.
 2. Number of road accidents in.

Continuous Variable: When a variable assumes infinite Uncountably number of values.

- Examples:**
1. Height
 2. Weight
 3. Sale, Profit etc.

Note: So far as the statistical analysis of the characteristic is concerned, we need to **convert qualitative information to quantitative information** by providing a **numeric description** to the given characteristic.

Method of Collecting Data

There are Two Ways of Collecting Data

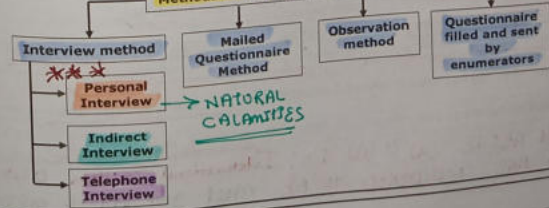
Primary data

The data which are **collected for the first time** by an investigator or agency are known as primary data

Secondary data

Secondary data can be defined as, data as being **already collected**, are used by a different person or agency.

Methods of collecting primary data



Method of Collecting Primary Data

1. Interview Method

Personal Interview: the investigator meets the respondents directly and collects the required information then and there from them.

Application: In case of a **Natural Calamity** like a super cyclone or an earthquake or an epidemic like plague, we may collect the necessary data much more quickly and accurately by applying this method.

Indirect Interview: Under this method, investigator collects the necessary information from the persons associated with the problems.

Application: if there are some practical problems in reaching the respondents directly, as in case of **Rail accident**.

Telephone Interview: Under this method, relevant information can be gathered by the researcher himself by contacting the interviewee over the phone.

Application: Quickest, Non-Expensive

2. Mailed Questionnaire Method

Questionnaire method comprises of framing a well-drafted and soundly-sequenced questionnaire covering all the important aspects of the problem under consideration and sending them to the respondents with pre-paid stamp after providing all the necessary guidelines for filling up the questionnaire.

Advantage: Wide Coverage & Non-Expensive

Disadvantages: Non-Responses

3. Observation Method

Under this method data are collected by direct observation or **using instrument**.
Example: Data on the height and weight of a group of students

4. Questionnaire filled and sent by enumerators

Questionnaire form of data collection is used for **larger enquiries** from the persons who are surveyed. Enumerators collect information directly by interviewing the persons having information. **Question are explained and hence data is collected.**

Method of Collecting Secondary Data

- (a) **International sources** like WHO, ILO, IMF, World Bank etc.
- (b) **Government sources** like Statistical Abstract by CSO, Indian Agricultural Statistics by the Ministry of Food and Agriculture and so on.
- (c) **Private and quasi-government sources** like ISI, ICAR, NCERT etc.
- (d) **Unpublished sources** of various research institutes, researchers etc.

AMONG ABOVE (a) & (b) i.e., International Sources, & Gov. Sources ARE CONSIDERED TO BE MOST IMPORTANT METHOD OF S. DATA.

Classification of Data

It may be defined as the **process of arranging data** on the basis of the **characteristic** under consideration into a **number of groups** or **classes** according to the **similarities** of the observations.

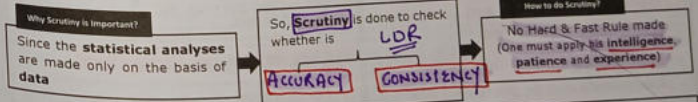
Objective

- a. It puts the data in a **neat, precise and condensed** form so that it is easily understood and interpreted.
- b. It makes **comparison possible** between various characteristics, if necessary, and thereby finding the association or the lack of it between them.
- c. **Statistical analysis** is possible only for the classified data.
- d. It **eliminates unnecessary details** and makes data more readily understandable.

Types of Classification of Data

Quantitative or Cardinal Classification	Qualitative or Ordinal Classification	Chronological or Temporal or Time series Classifying	Geographical or Spatial Classification
Basis of classification <i>Characteristics that can be measured in NUMBERS</i>	Basis of classification <i>Attributes that can not be Quantified.</i>	Basis of classification <i>It is difference of TIME PERIOD</i>	Basis of classification <i>Geographical location</i>
Example Height, Weight	Example Married or single ✓ Honest or dishonest ✓ Beautiful or ugly ✓ Hindu, Muslim, Sikh, Christian ✓	Example Years student 2002 → 1200 2007 → 1450 2009 → 1900	Example Place Population Araon → 31,205 Bhuban → 104099 Coo → 1458

Scrutiny of Data



How to Check Internal Consistency of Data?

CAN BE DONE ONLY :- WHEN NUMBERS OF RELATED SERIES ARE GIVEN

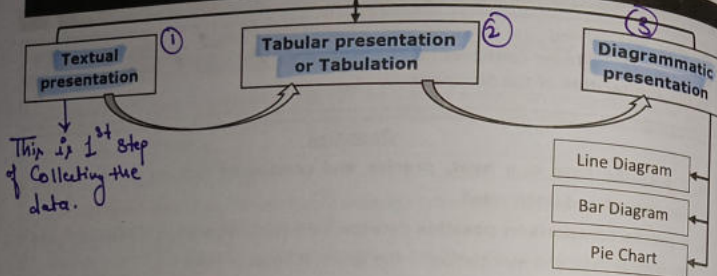
Note: If the data for population area and density for some places are given then we may verify internal consistency by examining following relation

DENSITY = AREA / POPULATION → THIS MUST HOLD TRUE

→ LHS = RHS



Mode of Presentation of Data



Textual Presentation

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

Example: The official report of an enquiry commission, Government Reports etc.

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.

In 2010, the number of workers belonging to the trade union was increased by twenty per cent as compared to 2009 of which four thousand and two hundred were male. The number of workers not belonging to trade union was nine hundred and fifty of which four hundred and fifty were females.

Merits

- The merit of this mode of presentation lies in its **simplicity** and even a **layman** can present and understand the data by this method.
- The observations with **exact magnitude** can be presented with the help of textual presentation.
- This type of presentation can be taken as the **first step** towards the other methods of presentation.

Demerits

- Dull
- Monotonous
- Does not facilitate Comparison**
- For manifold classification this method cannot be recommended.



Tabular Presentation

Tabulation may be defined as **logical** and **systematic** arrangement of statistical data in **rows** and **Columns**.

Objective of Tabular presentation

To simplify the complex data.	To clarify the objective of investigation	Economize space .	To facilitate comparison between rows & Column.	To depict trend a pattern of data
To facilitate statistical analysis .	To detect errors and omissions in the data.	It is must for Diagrammatic Presentation.		

Essential parts of Statistical Table

Box Head

Table No. - Table No. XXX Unit of measurement

Title: [Head note or prefatory Note]

Column 1 + Sub-column Caption

Stub Heading	Sub-Heads		Sub-Heads		Total (5)
	Column Head (1)	Column Head (2)	Column Head (3)	Column Head (4)	
Row 1					
Row 2		Body of Table			
Row 3					
Total					

Foot Note → (7)

Source Note: → (8)

Number and Date: No. Dc 1172-25/2/2006

(Most essential part of table)

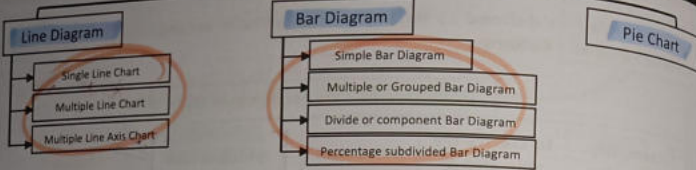
Brief of Some Important Parts of Table

Body of the Table: It is the most vital part of the table It contains the numerical information. It should be made as comprehensive as possible.

Source Notes: A note at the bottom of the table should always be given to indicate the primary source as well as the secondary source from where the data has been taken particularly. When there is more than one source.

Foot Notes and References: It is always placed at the bottom of the table. It is a statement which contains explanation of some specific items which cannot be understood by the reader from the title or captions and stubs.

Diagrammatic Presentation

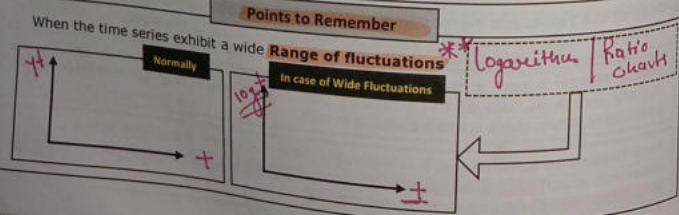


Line Diagram

It is the simplest of all the diagram. When the data vary over time, we draw a line diagram. We plot each pair of values of (t, y_t) , y_t representing the time series at the time point t in the t - y plane.

S.no	Types of Line Diagram	Diagram	Situation to Use
1.	Single Line Chart		WHEN WE WANT TO PRESENT THE DATA OF SINGLE TIME SERIES
2.	Multiple Line Chart		TWO OR MORE RELATED TIME SERIES. AND/OR DATA IS EXPRESSED IN SAME UNITS
3.	Multiple Axis Line Chart		TWO OR MORE RELATED TIME SERIES. AND/OR DATA IS EXPRESSED IN DIFFERENT UNITS

Points to Remember

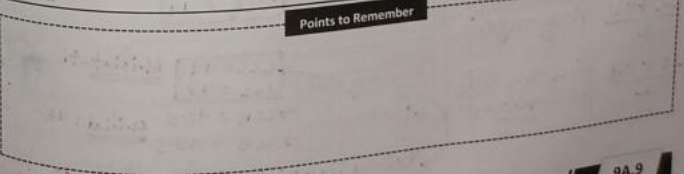


Bar Diagram

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

S.no	Types of Bar Diagram	Diagram	Situation to Use
1A.	Simple Vertical Bar Diagram		Quantitative data are given OR AND Time Series data is given.
1B.	Simple Horizontal Bar Diagram		Qualitative data is given OR AND DATA OVER WIDE SPACE.
2.	Multiple or Grouped Bar Diagram		COMPARISON OF TWO OR MORE RELATED VARIABLES (series).
3.	Divide or component Bar Diagram		COMPARING DIFF. COMPONENTS OF VARIABLE AND The relation of diff components to the total
4.	Percentage subdivided Bar Diagram		SAME THING as above Expressed in %

Points to Remember



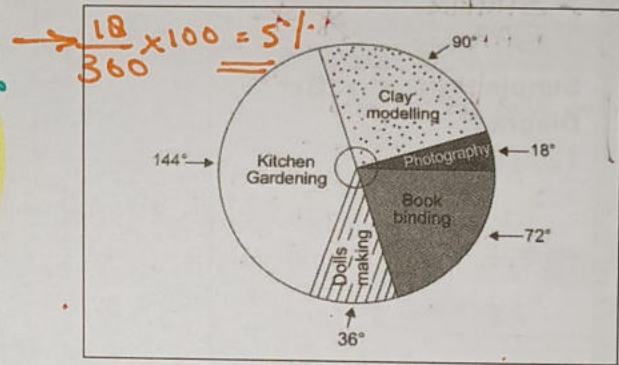
Pie Diagram

- A pie diagram is a **circular diagram** which represent the total value with its component. The area of circle represents the different parts.
- In a pie diagram data are expressed as **percentage**. Each Component is expressed as percentage of total value. The pie diagram is also known as **angular diagram**.
- The Surface area of Circle is **2π radians or 360°**

$$\text{Angle of the circle} = \frac{\text{Frequency of sector}}{\text{Total Frequency}} \times 360^\circ$$

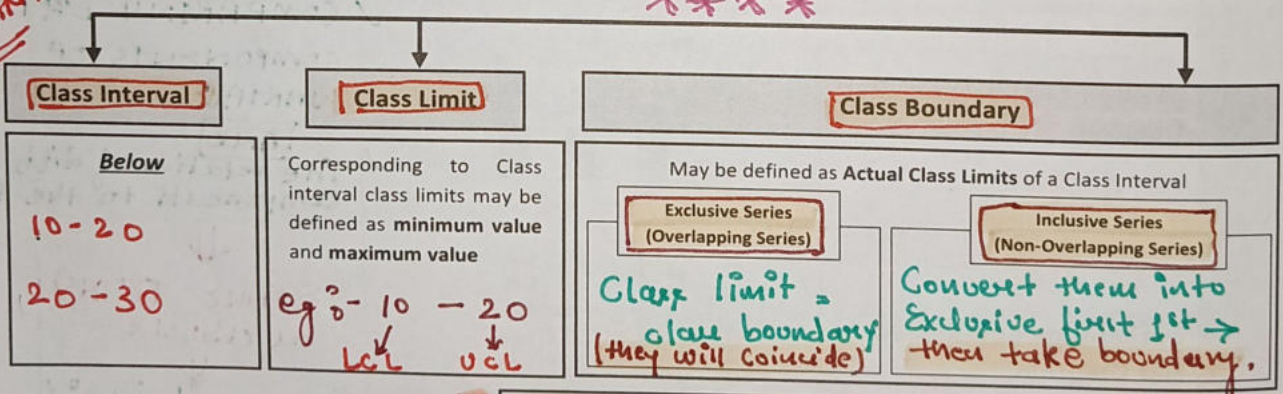
Example: 120 students of a college were asked to opt different work experience. The details of these options are as under.

Area of work	No of Students	Angle of Circle
Photography	6	$\frac{6}{120} \times 360 = 18^\circ$
Clay modelling	30	$\frac{30}{120} \times 360 = 90^\circ$
Kitchen Gardening	48	$= 144^\circ$
Doll Making	12	$= 36^\circ$
Book Printing	24	$= 72^\circ$
	120	$= 360^\circ$



Some Basics Terms that Students Should Know Before Studying Frequency Distribution

Imp most important



Mid-Point

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

$\frac{LCB + UCB}{2}$

Width ****

$UCB - LCB = \checkmark$

$UCL - LCL = \times$ WRONG

Example-1:

Weights in Kg	Freq
44-49	3
49-53	4

LCB = 44
UCB = 49

Example-2:

Weights in Kg	Freq
44-48	3
49-53	4

LCL = 44
UCL = 48

$\rightarrow LCB = 43.5$ M. Point = 46

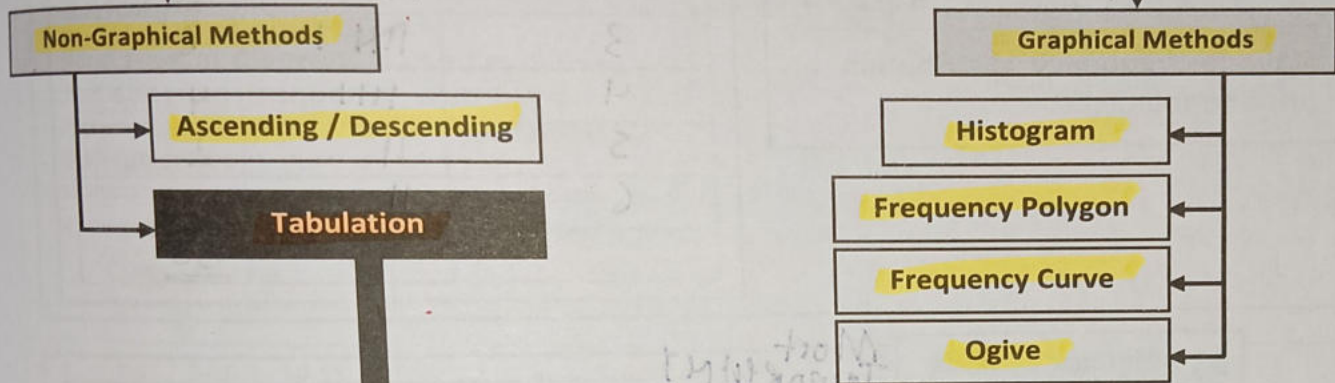
$\rightarrow UCB = 48.5$ M. Point = 46

Note: \checkmark LCB \Rightarrow lower limit to lower class limit
 \checkmark UCB \Rightarrow upper limit to upper class limit

Frequency Distribution

A frequency distribution is the arrangement of the **given data** in the form of a table showing **frequency** with which each variable occurs.

Mode of Presentation of Frequency Distribution



Terms Used to Draw Frequency Distribution Table (3 Most Common terms used)

Variable

Any **character** which can **vary** from **one individual** to another is called a variable or a variate.

Example: age, income, height, intelligence

Tally Marks

It is a method of keeping count in blocks of five.

Example

Age	Frequency	Tally
24	3	
35	8	

Frequency

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

Discrete Variable (Ungrouped Frequency)

Continuous Variable (Grouped Frequency)

It is a **Single Value** (Variable)
Example: No of Car accident in Delhi during 12 months

It is a **Group of Values** (Variables)
Example: Height of Students of St. Xavier's College for year 2004

Eg:- ✓ Marke of a student
 ✓ Distribution of shares
 ✓ Profit of a blue chip Co.
 ✓ Age of a person

In Case of Continuous Variable

Step-1: Find the Range: **NO NEED**

Step-2: Form the Number of Classes: **No. of class = $\frac{\text{Range}}{\text{class length}}$**

Step-3: Present Class & Class Interval in Table (Known as Frequency Distribution Table)

Step-4: Apply 'Tally Marks'

Step-5: Count tally marks write these in numbers in next column, known as frequency column.

WDR

Discrete Variable

Example: A review of the first 30 pages of a statistics book reveals the following printing mistakes:

0	1	3	3	2	5	6	0	1	0
4	1	1	0	2	3	2	5	0	4
2	3	2	2	3	3	4	6	1	4

Make a frequency distribution of printing mistakes.

Solution:

Frequency Distribution Table

Printing Mistake	Tally Marks	Frequency
0		5
1		5
2	1	6
3	1	6
4		4
5		2
6		2
		<u>30</u>

Continuous Variable

Most IMPORTANT

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

Make a frequency distribution of printing mistakes, Take Class length = 5

Frequency Distribution Table

Printing Mistake	Tally Marks	Frequency
43-48		3
49-53		4
54-58		5
59-63	11	7
64-68	1111	9
69-73	111	8
		<u>36</u>

Solution: Range = 73-44 = 29
 No. of class Range Interval = $\frac{\text{Range}}{\text{Class length}} = \frac{29}{5} = 5.8 \approx 6$

Some important terms associated with frequency Distribution (Do Learn it)

1 Frequency Density of Class Interval

Always lies b/w 0 to 1

$\frac{\text{frequency of class Interval}}{\text{class length}}$

2 Relative Frequency & Percentage Frequency of Class interval

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

$\frac{\text{class frequency}}{\text{total frequency}} \times 100$

Example

Weights in Kg	Frequency
44-48	3
49-53	4
54-58	5

RF: $\frac{3}{5} = 0.6$

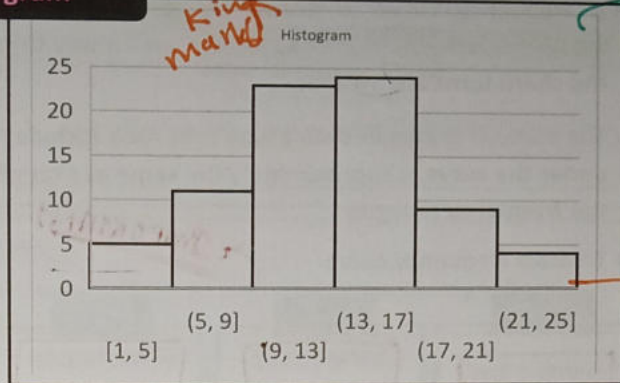
RF: $\frac{3}{12} = 0.25$

PF: $\frac{3}{12} \times 100 = 25\%$

Graphically Method to Present Frequency Distribution

Histogram

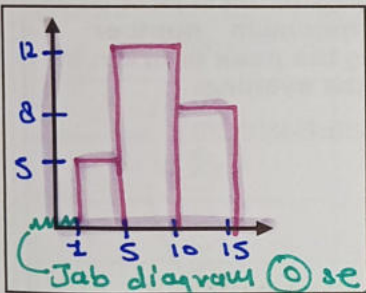
- A histogram is used to depict a **frequency distribution**.
- A Histogram is a graph containing a set of **class interval by its width** and the **frequency in each class interval by its height**.
- This type of diagram is used exclusively for showing frequency distributions of quantitative data that are **continuous in nature**.



① (Height) frequency (mode)
② (width) class interval

Histogram with Equal Class Interval

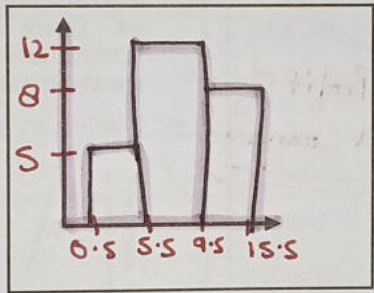
x	1-5	5-10	10-15
f	5	12	8



Histogram with Unequal Class Interval

x	1-5	6-9	10-15
f	5	12	8

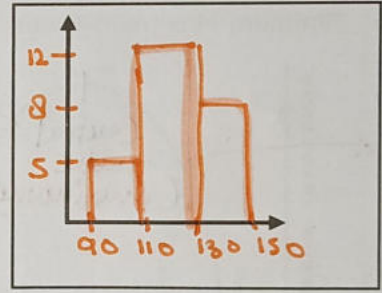
1st of all convert them Exclusive 1st.



Histogram When Mid-Points are Given

x	100	120	140
f	5	12	8

90-110 110-130 130-150



*** Histogram can be plotted where the series are Inclusive & Exclusive also
 $H = \frac{20}{2} = 10$

- ✓ Different Scale's may be taken on 2 axes (i.e., x-axis & y-axis).
- ✓ **King mark** (m) is used if class interval is **not starting from 0**.
- ✓ **In Histogram** → Height & width both matter. ✓
- ✓ **In Bar Diagram** → only Height matters. ✓

Frequency Polygon

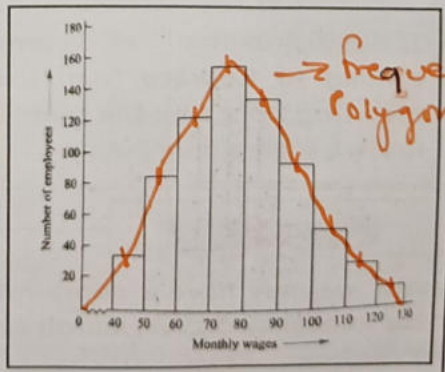
→ It is a limiting form of histogram.

- In order to draw frequency polygon **mid-point of the class interval** (mid-point of the top horizontal sides of rectangles in a histogram) and its **corresponding frequency** are plotted. The plotted points are joined successively by line segment which is known as Frequency Polygon.

Example: Draw Histogram and Frequency polygon

Monthly wages	No of Employees
40-50	36
50-60	87
60-70	121
70-80	154
80-90	133
90-100	95
100-110	50
110-120	30

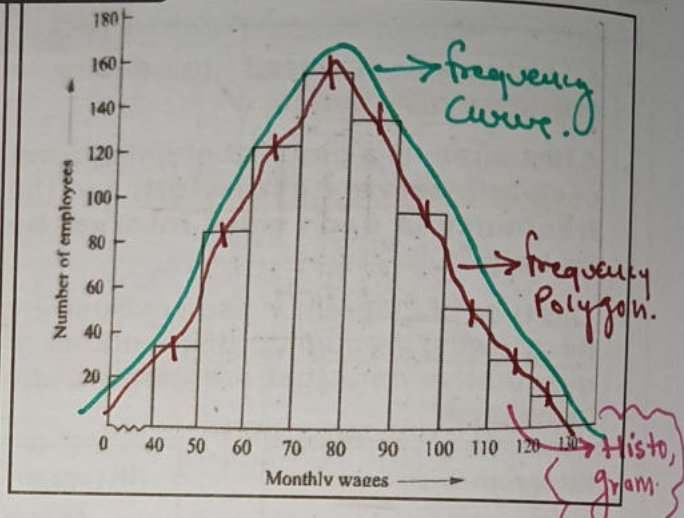
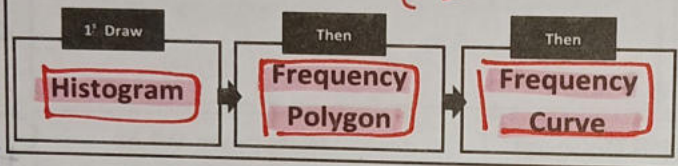
Solution:



Frequency Curve

limiting form of frequency polygon as well as histogram (and)

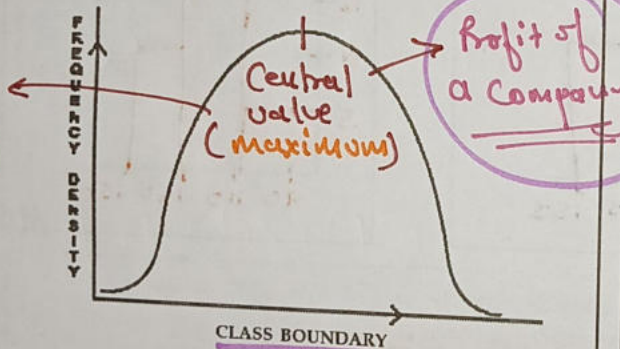
- A frequency curve is drawn by **smoothing** the **frequency polygon**. It is smoothed in such a way that the sharp turns are avoided
- The curve is drawn in such a way that Area included under the curve is approximately the same as that of the frequency polygon
- To draw frequency curve



Types of Frequency Curve

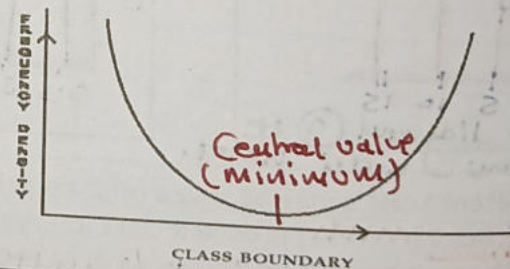
Bell-shaped curve

- Most of the **commonly used** distributions provide bell-shaped curve
- In case of **Profit** of a Company we should prefer this curve.
- Minimum-Maximum-Minimum



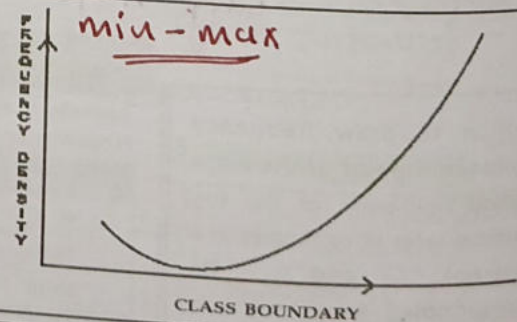
U-shaped curve

- 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.
- The distribution of **Kolkata** bound commuters belongs to this type of curve as there are maximum number of commuters during the **peak hours** in the morning and in the evening.
- Maximum-Minimum-Maximum



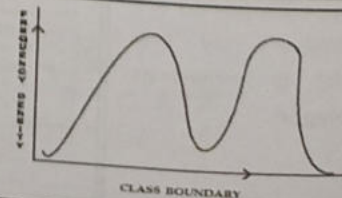
J-shaped

- The J-shaped curve starts with a minimum frequency and then gradually reaches its maximum
- The distribution of **commuters** coming to **Kolkata** from the **early morning** hour to **peak morning** hour follows such a distribution.



Mixed curve

Lastly, we may have a **combination** of these frequency curves, known as **mixed curve**.



Ogive Curve or Cumulative frequency Distribution

It is Obtained by plotting the graph between **Class Boundaries & Cumulative frequency**

Less than Ogive

Plotted Between

Cumulative Frequency with Upper Boundary

More than Ogive

Plotted Between

Cumulative Frequency with Lower Boundary

Assumption

It is assumed that class preceding the first class in the classification exist and its frequency is **Zero**

Example: Draw the less than and More than Ogive from following distribution.

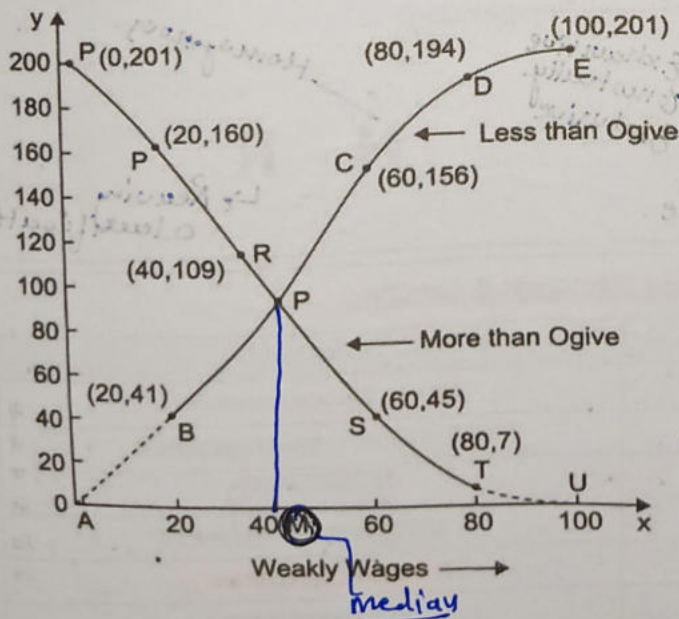
Weekly wages	0-20	20-40	40-60	60-80	80-100
No of Workers	41	51	64	38	7

Solution:

Weekly wages	Frequency	Cumulative Frequency	
		Less than	More than
0-20	41	41 ✓	201
20-40	51	92 ✓	160
40-60	64	156 ✓	109
60-80	38	194 ✓	45
80-100	7	201 ✓	7

201

0



Points to Remember

Draw a Straight line from point of intersection of two Ogive on X-axis, The Point of Intersection is **Median**

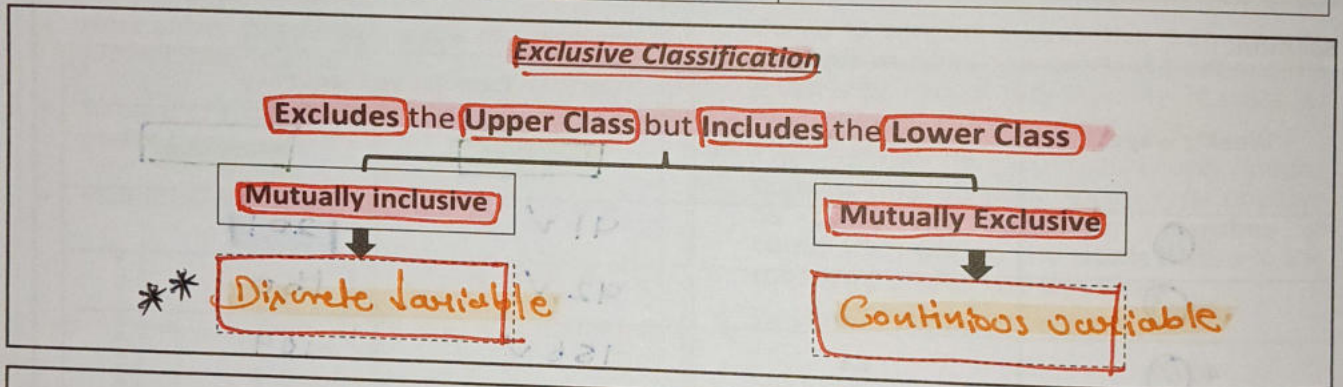
We Can even Say Ogive Help us to Find **Partition Values**

Median = $Q_2 = D_5 = P_{50}$

Note :- for particular class Boundary less than and more than C.F. Adds upto total frequency.

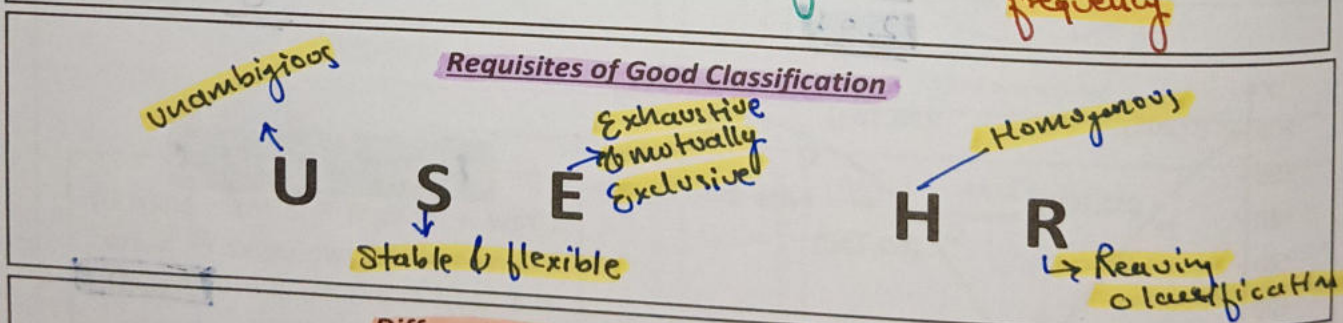
Important Highlights (Do Learn it)

S.no	Particulars	Appropriate Method
1.	Best Method to Present Data (Most Accurate)	TABULAR
2.	Most attractive (Hidden Trends)	DIAGRAMATIC
3.	Charts Uses Logarithm	RATIO CHARTS
4.	To Check Consistency & Accuracy	SCRUPINITY
5.	Internal Consistency can be Checked?	NUMBER OF <u>Releats series</u> we Given.



Formula to Compute Number of Class Interval

No. of class Interval $\Rightarrow 1 + 3.22 \log N$ — total of frequency



Difference Between Histogram & Bar Chart

Histogram	Bar Chart
Refers to <u>Graphical</u> Presentations	Refers to <u>Pictorial</u> Presentations
Distribution of <u>Continuous Variable</u>	Distribution of <u>Discrete Variable</u>
Quantitative Data	<u>Categorical Data</u>
Bars touches each other's	Bars Don't touch each other's
Bars can't be reordered	Bars can be reordered
Width of bars can or cannot be same	Width of bars is same



Other Useful Concepts

NOTE

frequency distribution

All of these

→ Arranging obs. in Increasing order

→ Arranging obs in terms of no. of groups

→ It relates to measurable character