

Chapter → 13

Statistical description of data

Statistics :] [Collection, analysis & presentation of data]

Data

Plural sense

Data qualitative or quantitative collected to do statistical analysis

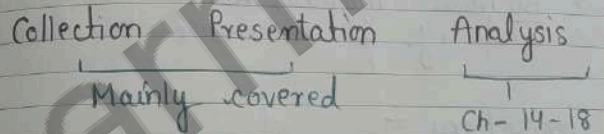
Singular sense

Used to draw statistical inferences

Inferences means conclusion reached based on evidence and reasoning

JUL 2023	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	
	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31						

Subject matter



History of statistics ÷

- Latin - Status
- Italian - Statista
- German - Statistic
- French - Statistique

> Kautilya's arthashastra

- Record of births and death
- Chandragupta's reign
- 4th Century BC

> Abu Fezal's Ain - I - Akbari

- Record on agriculture

> First census

- Egypt 300 B.C to 2000 B.C

JUL 2023	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	AUG 2023	
	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31								

Applications of statistics :-

- > Economics
- > Business management
- > Industry & commerce

Limitations :-

- Aggregate data :-

2, 7, 6, 4] Group

Individual data :-

2] Single

- > Relevant for aggregate data and not individual data

- Stats deal with only quantities
- But concerned with both quantity & quality

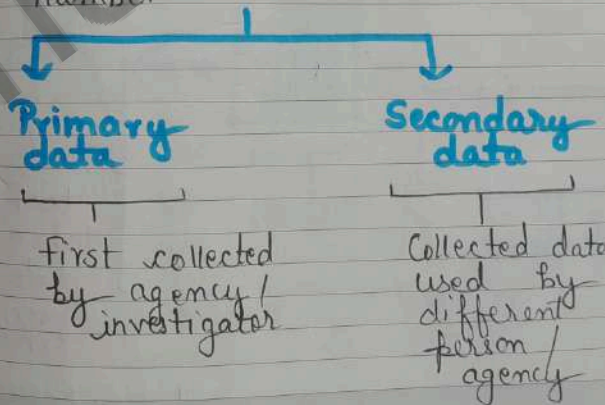
JUL 2023	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	
	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31							

Quantitative → Expressed in number only
 Qualitative → Not expressed in numbers

So mainly quantitative data is used and qualitative is converted into numbers

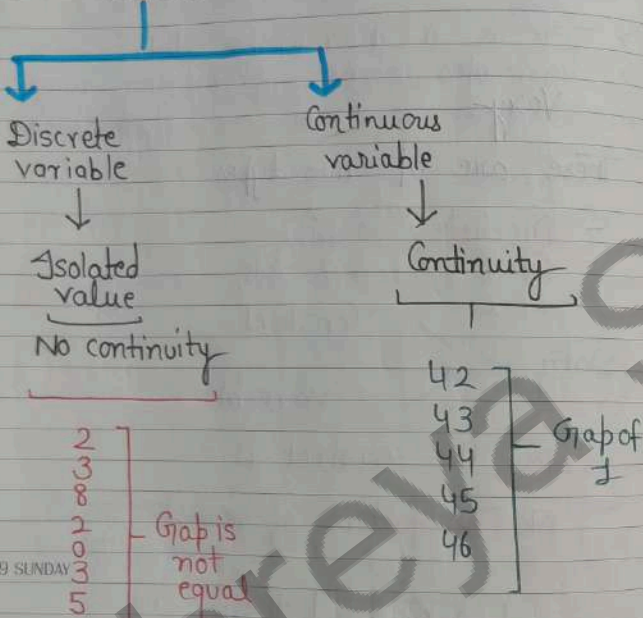
What is data :-

Qualitative information shown as number



AUG 2023	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	

Variable



Discrete variable

When variable assumes a finite or countably infinite isolated values

For eg: No of petals in a flower
No of Heads throwing coin for 2 times

Continuous variable

When there is continuity } A variable assumes any value from the given interval.
Can also be in decimals.

For eg:

Age of person in a camp.

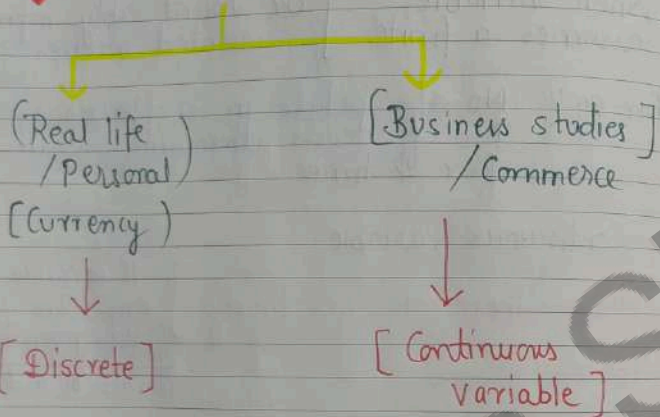
8.25, 9.37, 10.35, 10.85, 11.25

But continuous:-

Ex: Height, weight, sale, money

[Special]

MONEY



COLLECTION OF DATA

Will be of → Primary data
 [Bcz. secondary data is already collected]

JUL 2023	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	
	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31						

INTERVIEW METHODS :-

Personal / Direct Interview :-

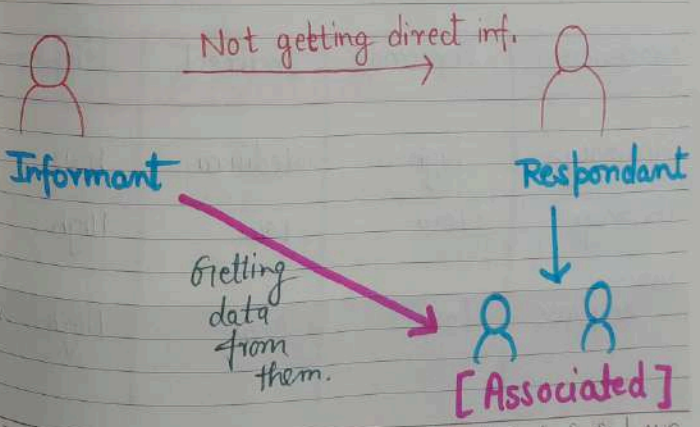


Informant

Respondant

Indirect interview :-

When we didn't get direct information from respondant we will get this from his associated person



Informant

Respondant

[Associated]

JUL 2023	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S		
	21	22	23	24	25	26	27	28	29	30	31											

Methods	Examples
★ Personal Interview	Natural calamity, door to door survey
★ Indirect Interview	Rail accident
★ Telephone Interview	By Calling / Messaging

Difference :-

Parameter	Personal	Indirect	Telephone
Accuracy	High	Medium	Low
Coverage	Low	Low	High
Non-Response	Low	Low	High

Mail Questionnaire method:-

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

Online form → Email

> Here coverage is wide but amount of non-response will be maximum.

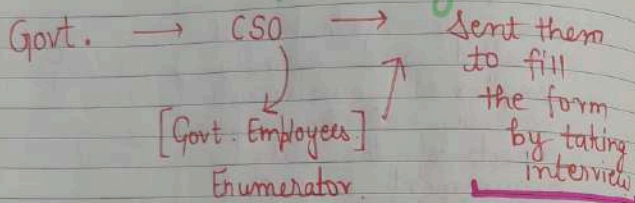
Observation method:-

> In this method data is collected by direct observation or using instrument

> Data on weight or height [Example]

★ > Although more accurate =
low coverage =
laborious method =

Questionnaire filled by enumerators



SOURCES OF SECONDARY DATA

★ International sources :-

- (WHO) World health Organization
- (IMF) International monetary Fund
- (ILO) International labour Organization
- World Bank

★ Government sources :-

In India

- Central statistics Office (CSO)
- National Sample survey office (NSSO)
- RBI, SEBI
- RERA
- IRDA

★ Private Or Quasi-govt. sources :-

- Indian Statistical Institute (ISI)
- Indian Council of Agriculture
- NCERT

Others

- Research papers and other unpublished sources.

Scrutiny of Data :-

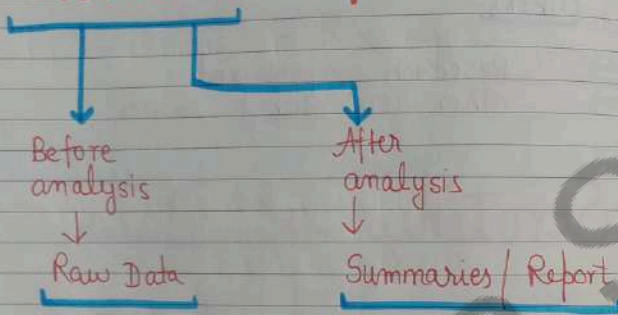
> Checking accuracy and consistency of data

> There is no rule for it, one must apply his intelligence, patience, experience while scrutinizing the given information

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

For eg: Age - 10 - 7 - 2000
DOB - 21] Not related

Presentation of Data :-



Presentation - Before analysis.

> Classification or organisation of Data

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

JUL 2023	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	
	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31							

> Objectives :-

- Neat, Precise, Condensed Data ^{No unnecessary details}
- Facilitate comparison
- Statistical Analysis

★ Classification

- Chronological / temporal / time series data.

Data arranged based on time
Eg: Revenues YOY
Year on year

- Geographical or spatial series data.

Arrangement based on regions

Eg: Country wise revenue of a company

JUL 2023	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	AUG 2023										
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	

On the basis of quality :-

- Qualitative or ordinal data :

Based on some attribute

Eg : Olympic medals by a company

- Qualitative or cardinal data :-

Based on some variable

2, 2, 3, 3, 3, 5, 4, 5, 2, 4
V. F
2 3
3 3
4 2
5 2

Classification	Temporal	- Time
	Spatial	- Region
	Ordinal	- Not Quantitative (Qualitative)
	Cardinal	- Quantitative

JUL	M	T	W	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	
2023	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31						

MODE OF PRESENTATION :-

- Textual Presentation :-

There is a company which has 30 male and 40 female. 20 are married and all

↓
This is textual presentation

- Tabular presentation :-

- When data is shown in the form of Table

- It is preferred over textual form bcz of the below reasons

- Useful in easy comparison.
- Complicated data can be presented
- Table is must to create a digram
- No analysis possible without table

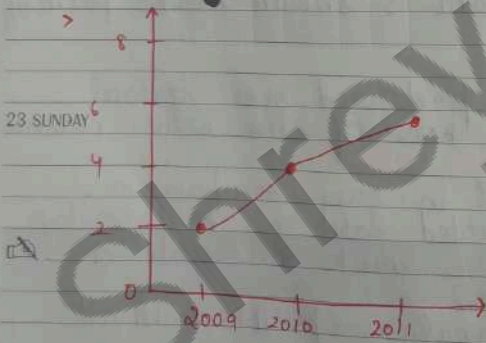
W	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S
21	22	23	24	25	26	27	28	29	30	31										

Diagrammatic presentation

- > Can be helpful for layman [without having much knowledge of numbers]
- > Hidden trend can be traced
- > But table is more accurate than diagrams

Types of diagram :-

Line diagram :-



- > Plotting points in graph and join them to make a line
- > Used generally for time series (variable y is plotted against t)
- > For wide fluctuation log chart or ratio chart is used (log y is plotted against t)
- > For two or more series of same unit - multiple line charts used.

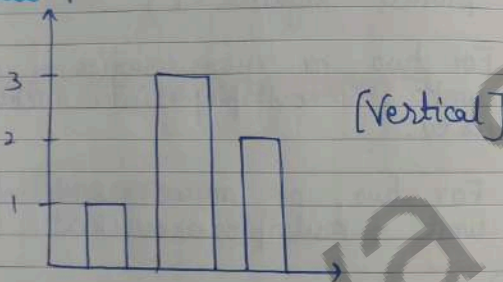
- > For two or more series of distinct unit - multiple axes chart is used

Bar diagram

- > Bar is a rectangle of same width and of varying length drawn horizontally or vertically
- > For comparable series - multiple or grouped bar diagrams can be used

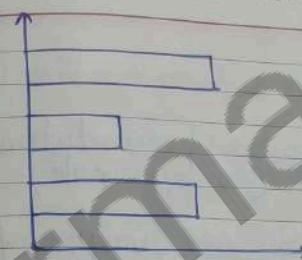
- > For data divided into multiple components - subdivided or component bar diagrams
- > For relative comparison to whole percentage bar diagrams or divided bar diagrams

Overview :-



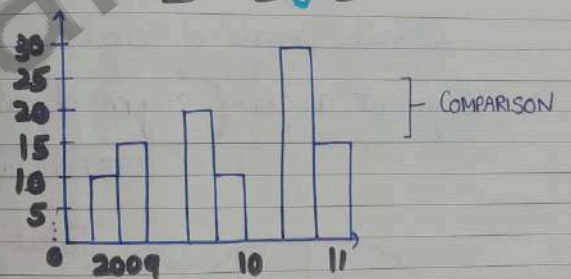
Vertical bar diagram is used for time series data or quantitative data / cardinal data

⇒ Historical bar diagram is used for geographical / spatial data / qualitative or ordinal data

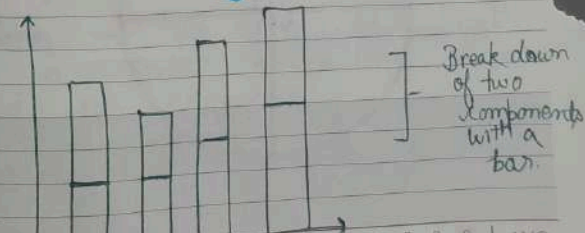


Horizontal bar diagram

Grouped bar diagram



Component bar diagram :-

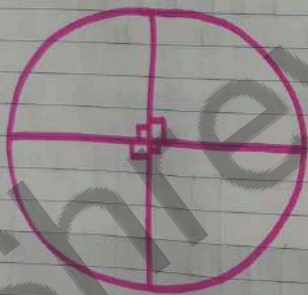


Pie chart :-

- > Used for circular presentation of relative data (% of whole)
- > Summation of all values of all components / segments equated to 360 Degree

(Total angle of circle)

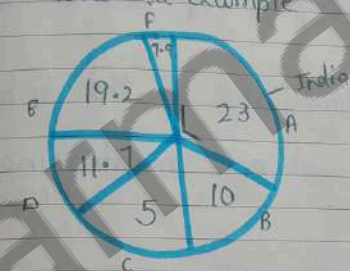
> Segment angle = $\frac{\text{Segment value} \times 360}{\text{Total value}}$



Total = 360°
4 equal part = $\frac{360}{4}$
= 90°

Just take an example =

Total export = 768



Segment angle = $\frac{23}{76.8} \times 360^\circ$
Sector angle = 107.81° = A Ex.

About table :-

Caption - The (lower) upper part of a table describing columns and sub columns.

Box head - Entire upper part of table.

- 3. **Stub** : Left part of table describing rows.
- 4. **Body** : Main part of the table.
- 5. **Footnote** : A note mentioning source of data.

Gender Unit	Member of Trade union		Not member of trade union		TOTAL	
	Male	Female	Male	Female	Male	Female
	% No	% No	% No	% No	% No	% No
2009						
2010						

Caption

30 SUNDAY

Stub

for Annual report of

Stub.

Footnote.

FREQUENCY DISTRIBUTION

Frequency : Number of times a particular observation is repeated

> Frequency distribution is a table which contains observations or class intervals in one column and corresponding frequency in the other.

Eg. 2, 2, 3, 0, 0, 1, 2, 3, 4, 4
[ungrouped]

Observation	Frequency
Ob. 1 = 0	2
Ob. 2 = 1	1
Ob. 3 = 2	3
Ob. 4 = 3	2
Ob. 5 = 4	2

- Individual values

Frequency distribution.



NOTES

[Grouped Frequency Distribution

When values are not limited =

Take grouped frequency distribution

Your choice

Within [class intervals]

↓
Class intervals.

Here you can take tally marks.

CLASS LIMIT

For a class interval CL is the minimum and maximum value the class interval may contain

> Minimum: LCL (Lower)

> Maximum: UCL (Upper)

Eg: From discrete series :-

Class Interval	Frequency	LCL	UCL
10-19	10	10	19
20-29	5	20	29
30-39	8	30	39

Non overlapping
OR
Mutually Inclusive

In case of continuous series

Class Interval	Frequency
10-20	
20-30	-
30-40	
40-50	

Where to take 20?

Ans: Where 20 is UCL

[Overlapping or mutually exclusive]

Previously

Classification of grouped frequency

- > Mutually exclusive
- > Mutually inclusive

Mutually exclusive :-

- > Also known as overlapping classification
- > Here UCL of an interval and LCL of next interval are same.
- > This is usually applicable for continuous variable.

- > An observation which is equivalent to common class limit is excluded from the class interval where it is UCL and taken in the class where it is LCL

Class	LCL	UCL
10-20	10	20
20-30	20	30

Mutually inclusive :-

- > Also known as non-overlapping classification
- > There is no common class limits between two intervals.
- > This is usually applicable to discrete variable
- > All observation including UCL and LCL will be taken in the same class interval as there is no confusion.

Class	LCL	UCL
10-19	10	19
20-29	20	29
30-39	30	39

Class Boundary

- > For mutually exclusive / overlapping :-

- Class Boundary = Class Limit
- LCL = LCB
- UCL = UCB

- > For mutually inclusive / Non-overlapping :-

- Mid of the two class limits

$$- LCB = LCL - \frac{D}{2}$$

$$- LCB = UCL + \frac{D}{2} \dots \dots \text{Class interval difference}$$

MID POINT

$$1. \rightarrow \frac{LCI + UCL}{2}$$

OR

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

Answer will be same

Class Length

$$= UCB - LCB$$

06 SUNDAY

For eg: UCB = 29.5
LCB = 19.5

so Class Length = $29.5 - 19.5 = 10$

CUMULATIVE FREQUENCY

Class Interval	Frequency	UCB	More than C.F	Less than C.F	Total A+B
44-48	3	48.5	33	3	36
49-53	4	53.5	29	7	36
54-58	5	58.5	24	12	36
59-63	7	63.5	17	19	36
64-68	9	68.5	18	28	36
69-73	8	73.5	0	36	36

Class Freq. density

$$= \frac{\text{Frequency}}{\text{Class length}}$$

Relative Frequency :-

> for a class interval, relative frequency

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

Weight	F	Relative f	Percentage F
44-48	3	3/36	8.333%
49-53	4	4/36	11.111%

for %f

x Relative f by 100

i.e = $\frac{\text{Class } f}{\text{Total } f} \times 100$

HISTOGRAM :-

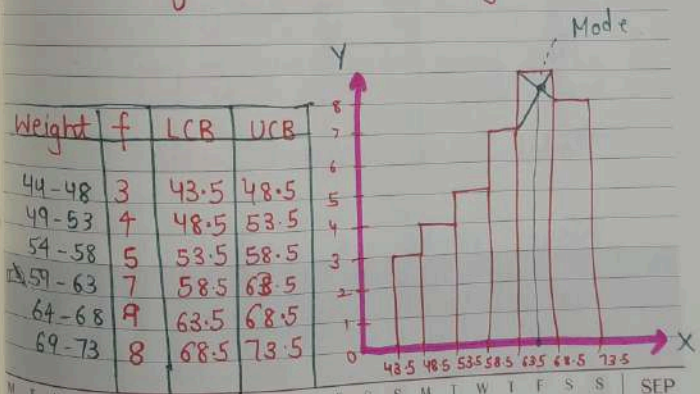
→ It is a convenient way to express FD

→ Comparison between frequency of two different classes possible

ALIG 2023	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21

→ It is useful to calculate mode also
steps to create :-

- Convert CL into CB and plot in X axis
- Form rectangles taking class intervals as base X axis.
- And frequency as length Y axis
- Use frequency density in case of uneven length.



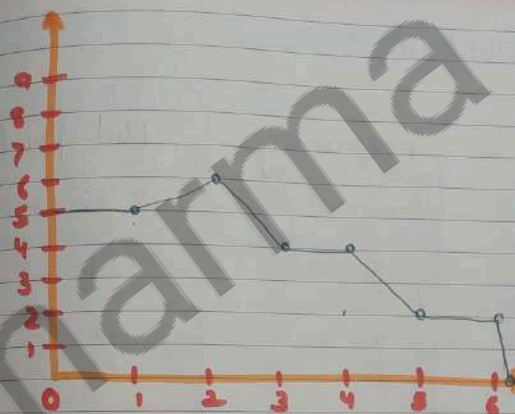
SEP 2023	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	

In case of uneven length..
Use frequency density ...

FREQUENCY POLYGON

Observation	Frequency
0	5
1	5
2	6
3	4
4	4
5	2
6	2
Total	30

Now make of this a graph



- > Usually preferable for the ungrouped frequency distribution.
- > Can be used for group also but if class lengths are even.
- > Steps to create =

X axis = Class value
Y axis = Frequency

Join all plotted points to make line segment which eventually will become a polygon.

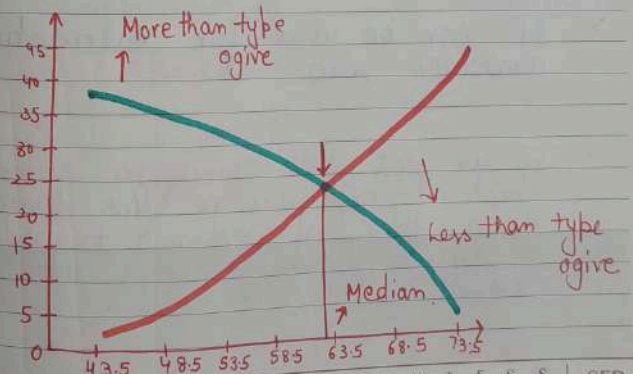
If grouped :-

Weight	Frequency	Mid point
44 - 48	3	46
49 - 53	4	51
54 - 58	5	56
59 - 63	7	61
64 - 68	6	66
69 - 73	8	71



Ogives / Cumulative Frequency graph :-

Class Boundary	Less than type C.F	More than type C.F
43.5	0	36
48.5	3	33
53.5	7	29
58.5	12	24
63.5	19	17
68.5	28	8
73.5	36	0



Steps ÷

- > Create a table where cumulative frequency is mapped against each class boundary and make a curve by plotting and joining points by line segments.

[Curve is called Ogive]

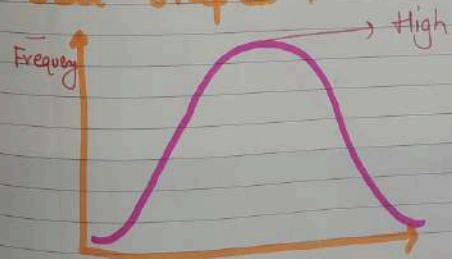
- > This graph can be made by both type of CF and called as Less than Ogive or more than Ogive.

- > It can be used for calculating quartiles also.

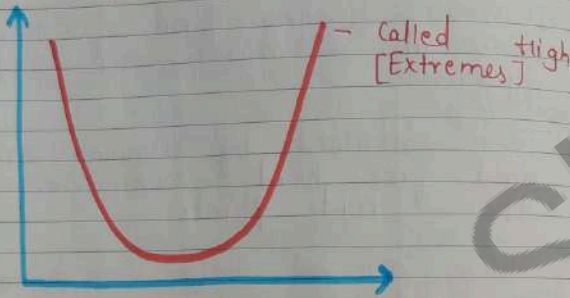
- > If we plot both ogives in same graph, perpendicular line drawn from their intersection towards x axis is cutting axis at Median.

FREQUENCY CURVE ÷

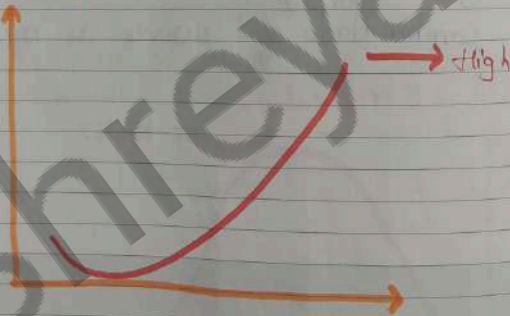
- > It is a limiting form of area diagram (Histogram) or frequency polygon.
- > It is obtained by drawing smooth and free hand curve through the mid points.
- > These are of below four types ÷
 - Bell shaped
 - U shaped
 - J shaped
 - Combination of curves as mixed curve

Bell shaped ÷

U shaped ÷

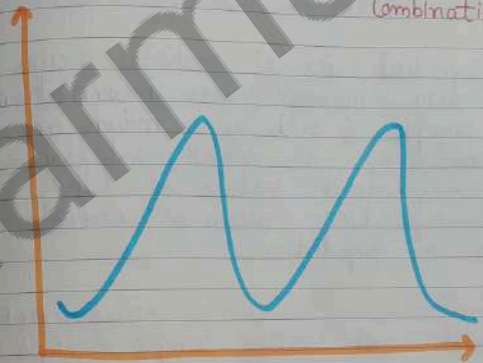


J shaped ÷



Mixed Curve ÷

Combination of waves ÷



Descriptive statistics ÷

→ When a statistic is used only to describe scores in a sample (and not used to make inferences about populations) that is a descriptive use.

→ For eg: a teacher may compute the mean test score for her class. She simply uses this number to think about how well her students did.

19

SATURDAY
DAY 231-134

UNIK PIPE FITTINGS + UNIK VALVES
Largest Pipe Fittings Group of India

AUGUST
2023

WEEK 30

Inferential statistics :-

- > Inferential use of statistics occurs when a data analyst uses information from a sample (such as a mean or correlation) to make inferences or guesses about values of the corresponding mean or correlation in a population.
- > Statistical interference only works well with large sample sizes.
- > For example, in a political poll, a polling organisation obtains the percent of people who state an intention to vote for candidate X for a sample of votes, perhaps

20 SUNDAY

selected by random digit telephone dialing from the population of all registered voters.

AUG	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S
2023	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	

AUGUST
2023

WEEK 31

UNIK PIPE FITTINGS + UNIK VALVES
Largest Pipe Fittings Group of India

21

MONDAY
DAY 233-132

Simple Vs. Manifold Classification

Simple classification :-

- > When based on only one attribute, the given data is classified into two classes, which is known as simple class classification.
- > For eg: When the population is divided into literate and illiterate, it is a simple classification.

Manifold classification :-

- > When based on more than one attribute, the given data is classified into different classes, and then sub-divided into more sub-classes, which is known as manifold classification.
- > For eg: when the population is divided into literate and illiterate then sub-divided into male & female, and further sub divided into married or unmarried.

AUG	M	T	W	T	F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S
2023	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21

22

TUESDAY
DAY 234-131

UNIK PIPE FITTINGS
UNIK+ VALVES
Largest Pipe Fittings Group of India

AUGUST
2023

WEEK 34

Ideographs / Pictograms :-

- A symbol that represents an idea or a thing, rather than the sounds of a word, is called an ideograph.
- A smiley face emoji is an ideograph that represents happiness.
- Many street signs ideographs, meant to convey a specific meaning without using any words.

Qualitative data types :-

Nominal data :-

- > Nominal data are used to label variables without any quantitative value.

→ Common examples

- Color
- Names of people
- Nationalities

AUG | M | T | W | T | F | S | S | M | T | W | T | F | S | S

AUGUST
2023

WEEK 34

UNIK PIPE FITTINGS
UNIK+ VALVES
Largest Pipe Fittings Group of India

23

WEDNESDAY
DAY 235-130

Ordinal data :-

- > Ordinal data comes from order.
- > Ordinal scales are often used for measures of satisfaction, happiness and so on.
- > Example → high - low - medium
→ strong - weak - average etc.

Research Data :-

Placed into two broad categories

Quantitative data :-

- Used when a researcher is trying to quantify a problem. It addresses the "what" or "how many" aspects of a research question.
- It is data that can either be counted or compared on a numeric scale.

For eg: it could be the no. of 1st year students at Macalester

M | T | W | T | F | S | S | M | T | W | T | F | S | S | SEP

, or the ratings on a scale of 1-4 of the quality of food served at Cafe Blue. This data is usually gathered using instruments, such as a questionnaire which includes a rating scale or a thermometer to collect whether data.

→ Statistical analysis software such as SPSS is often used to analyze quantitative data.

Qualitative data :

→ Qualitative data describes qualities or characteristics.

→ It is collected using questionnaire, interviews or observations and frequently appears in narrative form.

For eg: it could be notes taken during a focus on the quality of food or responses from an open ended questionnaire.

→ Qualitative data may be difficult to precisely measure and analyze.

→ The data may be in the form of descriptive words that can be examined for patterns or meaning, sometimes through the use of coding.

→ Coding allows the researcher to categorize qualitative data to identify themes that correspond with the research questions and to perform quantitative analysis.

Sturges' Rule : [20th century]

A method of choosing the optimum no. of bins in a histogram that minimize the potential for these pitfalls.

$$k = 1 + 3.32 \log n$$

↓
No. of bins

↓
No. of observations in a data set