12 May 2024 23:56

Statistical Description of Data

Statistics is originated from

Latin word – "Status"
Italian word – "Statista"
German word – "Statistik"
French word – "Statistique"

Definition of statistics

Singular sense ⇓ Scientific methods ⇓
→ Collection of Data
→ Classification of Data
→ Presentation
→ Analysing the Data
→ Interpretation

Plural sense ⇓ Numerical facts collected systematically with a definite purpose

Application of statistics

11 Application of Statistics

→ Economics
→ Business management
→ Commerce & Industry

Limitation of Stats

→ Deals in aggregates only
→ Does not deal in qualitative data
→ Not always true
→ Expert consultation required
→ Stats can be misused

Data

Information about some particular character

Quantitative character
⇓
"Variable"
⇓
Can be measured

Descrete
⇓
Particular number

Continous
⇓
Decimals & Fractions

Qualitative character
⇓
"Attributes"
⇓
Can not be measured

Eg → Colour of flower
Eg → Nationality of a person
Eg → Community of a person

Particular number

- No. of Accidents
- Goals in a match
- Salary

Decimals & fractions

- Weight
- Height
- Age
- Temperature.

g → community of a person

Collection of Data

If Data is Primary

"This is collected for first time"

Methods

#1] Interview method

- Direct (Personal interview)
  - Natural calamity
    - Cyclone
    - Earthquake
- Indirect
  - Death accident
- Telephonic (fastest)

#2] Questionaire method

If Data is secondary

Public / Existing information used by others

Methods

#1] International sources

- WHO, IMF, World Bank

#2] Govt sources

- Surveys from Different ministries

#3] Private sources

#4] Unpublish sources

Mailed questionair
↓
→ wide Area
→ maximum non Responses

Enumerator method
↓
→ for uneducated informants

#3] Observation Method
↓
used for research to gather data about people, objects, events & behaviours etc.

→ Accurate data
→ Time consuming

Scrutiny of Data
↓
Verification
↙ ↘
Accuracy of Data | consistency of Data

Classification of Data

1) Chronological or Temporal - on the Basis of time

eg

Year	2010	2011	2012
Sales	100	120	150

Eg

Year	2010	2011	2012
Sales	100	120	150

2> Geographical or Spatial – on the basis of Location (Place)

Eg

States	UP	Punjab	Bihar	Goa
Rice production	100	80	120	60

3> Qualitative (or ordinal)

Eg

Caste	Rajput	Brahman	Siya	Sunni
Population	1000	1200	600	400

4> Quantitative or Cardinal

Eg

Marks	1	2	3	4	5
No. of Students	6	8	4	10	20

Presentation of Data

- Textual
  - Paragraph
    - worst method
- Tabular Presentation
  - Table
- Diagrammatic Presentation
  - Line Diagram
  - Bar Diagram
  - Pie chart

worst method

Table
↓
→ Best method
→ most accurate

Line Diagram, Bar Diagram, Pie chart
↓
most attractive
(can notice hidden trend in data)

Textual method

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.

Table

1> Table Number
2> Title (which explains content of tables)
3> Headnote (given in brackets information about the units)
4> Caption - Describes column & subcolumns
5> Stubs → Row Head
6> Body of Table - Numeric information
7> Source of Data
8> Footnote - specific feature which is not explained

Box Head
↓
Upper part of Table
includes column & subcolumn number

Table No 25
Number of Teachers

lable No 25

Number of Teachers

Subjects	Class 11th		Class 12th		Total
	male	Female	male	Female	
Maths	3	1	2	2	8
A/c	2	2	3	2	9
Eco.	1	1	2	1	5
English	1	2	2	4	9
Total	7	6	9	9	31

Source :

Footnote : new joinees one not recorded Here.

Diagrammatic method

⇒ 1> Line Diagram → can be used for time series

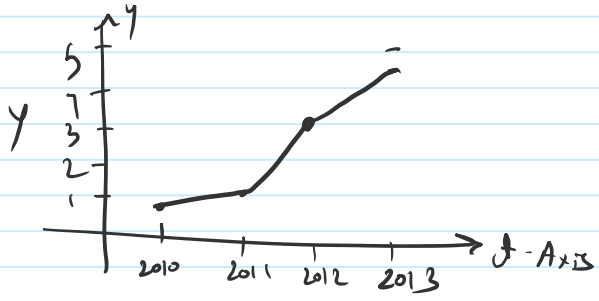
↓

x-y Plane

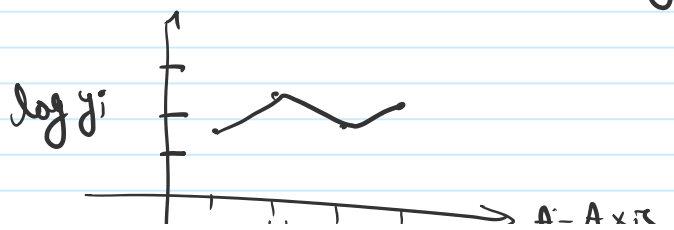
" There is no x-Axis, there is x Axis & y- Axis "

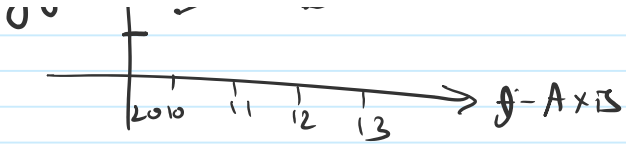
↓

one Dimensional



Ratio Chart : when time series have fluctuations we use Ratio chart (use Log yi)

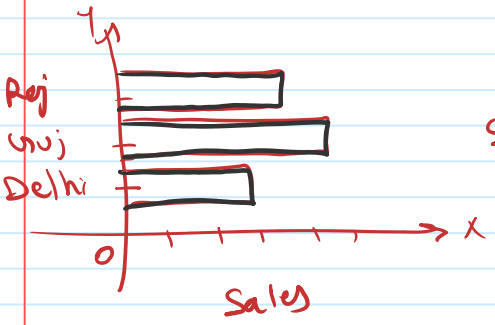


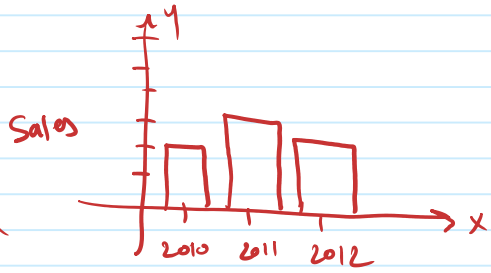


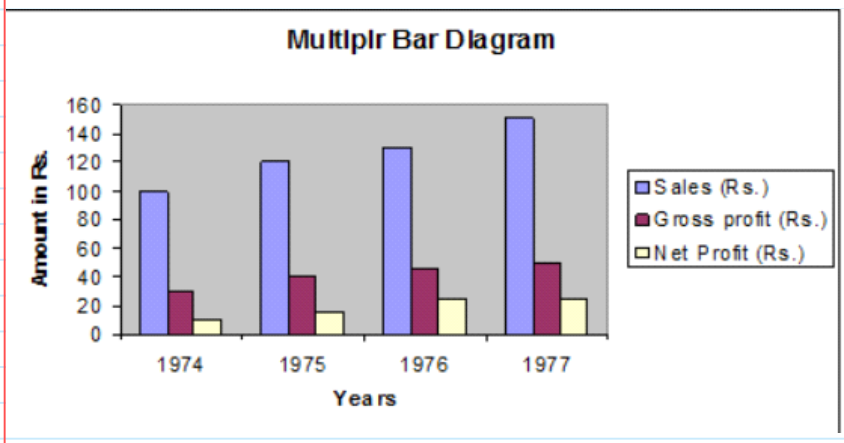
⇒ 2] Bar Diagram → Rectangle with Equal width
→ one Dimensional

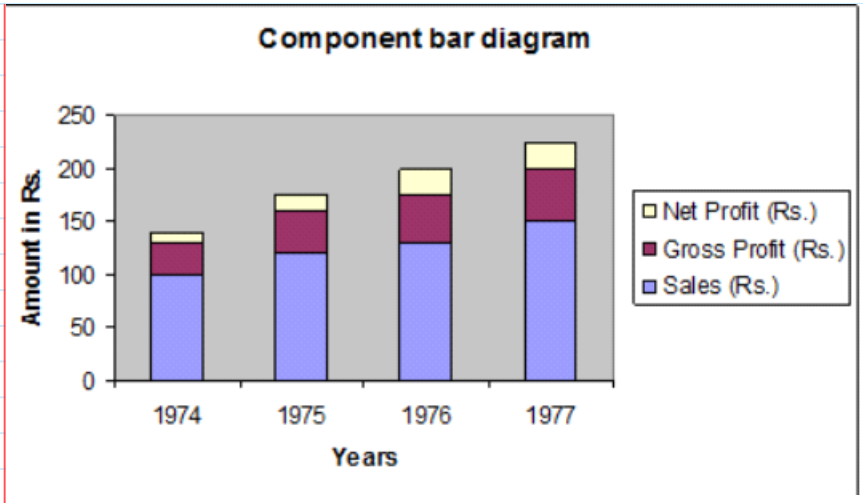
Horizontal
⇓
Qualitative
&
vary over space

vertical
⇓
Quantitative
&
vary over time



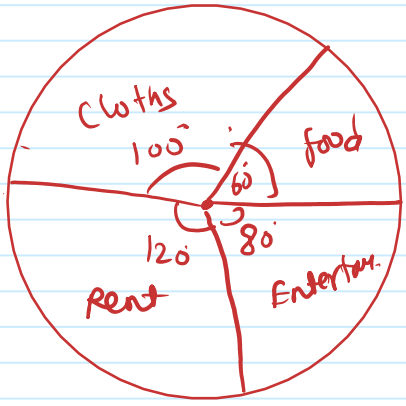






⇒ 3] Pie chart

Expense		central angle
food	₹ 3000	$\frac{3000}{18000} \times 360^\circ = 60^\circ$
Cloths	₹ 5000	$\frac{5000}{18000} \times 360^\circ = 100^\circ$
Rent	₹ 6000	$\frac{6000}{18000} \times 360^\circ = 120^\circ$
Entertainment	₹ 4000	$\frac{4000}{18000} \times 360^\circ = 80^\circ$
	18000	360°



Frequency Distribution

⇓

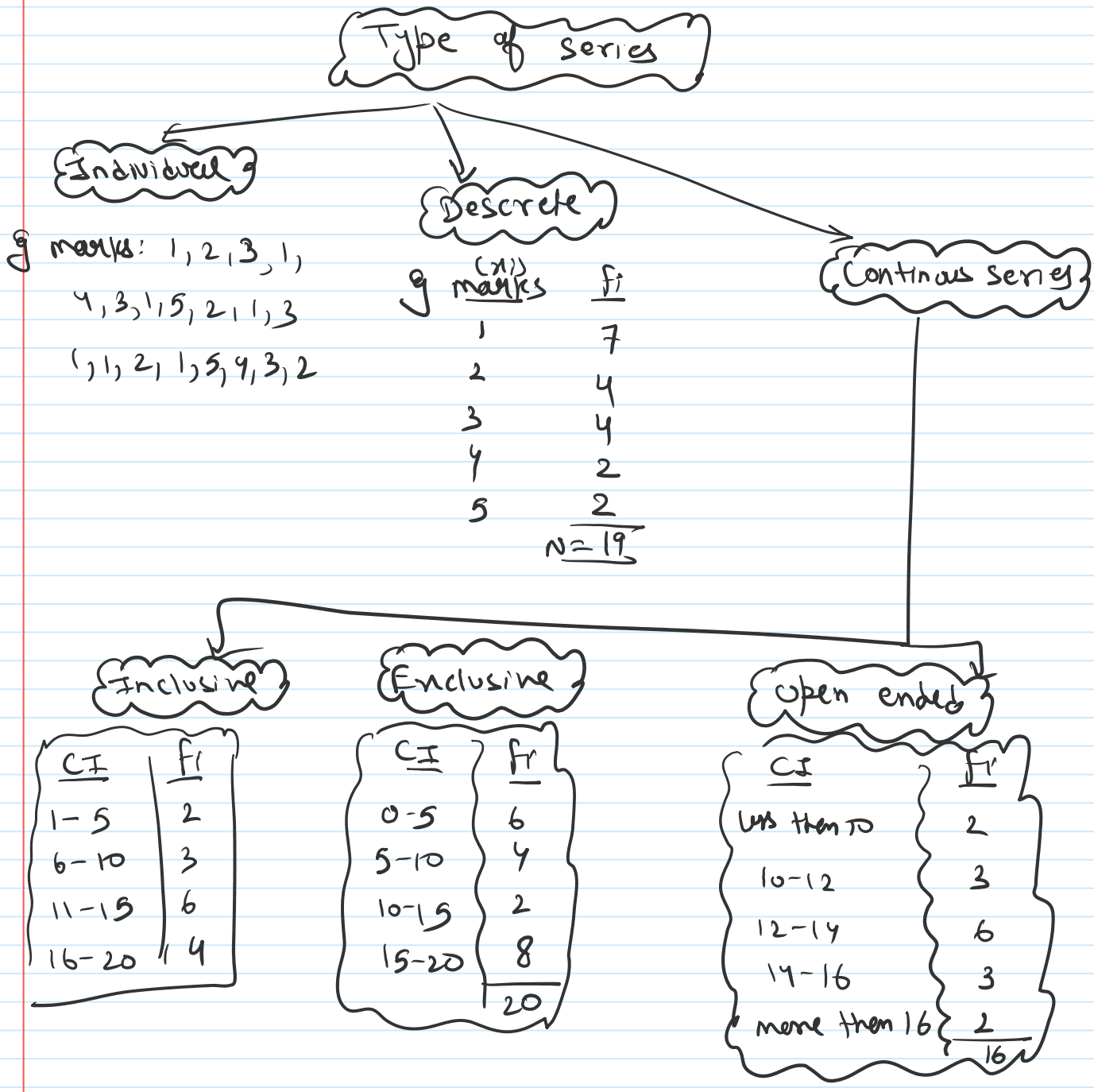
Tabular presentation of statistical data

No. of Accidents in last 15 Days

1, 2, 1, 1, 1, 2, 3, 3, 2, 1, 1, 1, 2, 3, 2

No. of Accidents	Tally Bar	Frequency
1	𝍸 ||	7

No of Accidents	Tally Bar	Frequency
1									7
2							5
3					3
		N=15



g

CI	fi
1-5	6
6-10	8
11-15	4

convert

CI	fi
0.5 - 5.5	6
5.5 - 10.5	8
10.5 - 15.5	4

1 - 5
- Lower class Limit (LCL)
- upper class Limit (UCL)

0.5 - 5.5
- Lower class boundary
- upper class boundary

Mid point of LCL & UCL is known as class mark

Frequency Density of a class interval = $\frac{\text{Frequency of class interval}}{\text{class length}}$

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

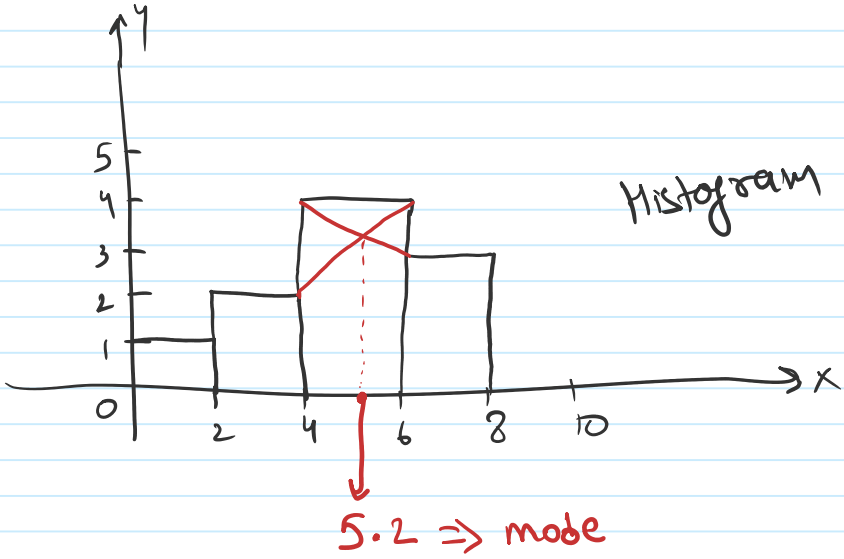
CI	fi	Freq Density	Relative Frequency	% Frequency
0-5	2	$\frac{2}{5} = 0.4$	$\frac{2}{20} = 0.1$	10%
5-10	12	$\frac{12}{5} = 2.4$	$\frac{12}{20} = 0.6$	60%
10-15	6	$\frac{6}{5} = 1.2$	$\frac{6}{20} = 0.3$	30%
	20		1	100

Graphical representation of a frequency curve

- Histogram → 2D
- Frequency Polygon → 1-D
- ogive → 1-D

Histogram

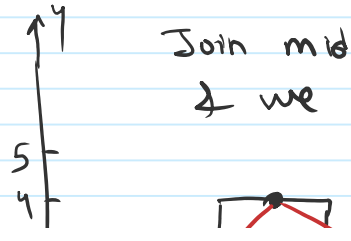
CI	FI
0-2	1
2-4	2
4-6	4
6-8	3



we can find mode graphically using Histogram

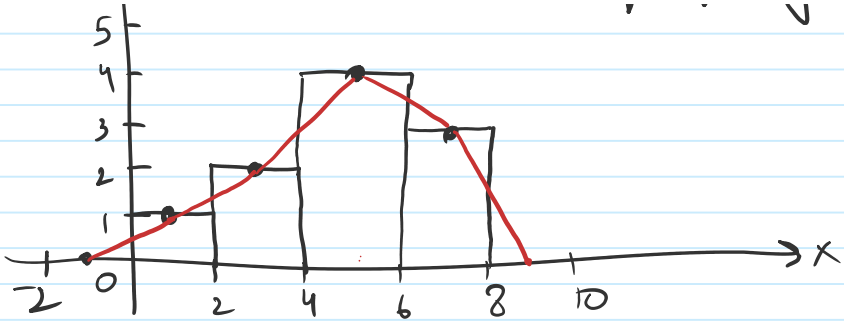
Frequency Polygon

CI	FI
0-2	1
2-4	2



Join midpoint of each Histogram & we obtain frequency polygon

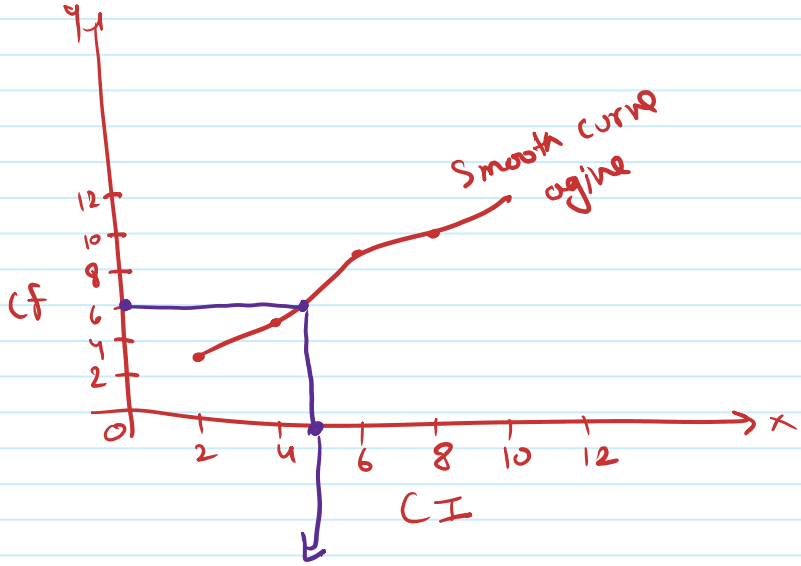
0-2	1
2-4	2
4-6	4
6-8	3



ogive

CI	fi	cf
0-2	3	3
2-4	2	5
4-6	4	9
6-8	1	10
8-10	2	12
	12	

cumulative frequency



$N = 12$

$\frac{N}{2} = 6$

5 = median

"we can find median graphically using ogive"

Frequency curve

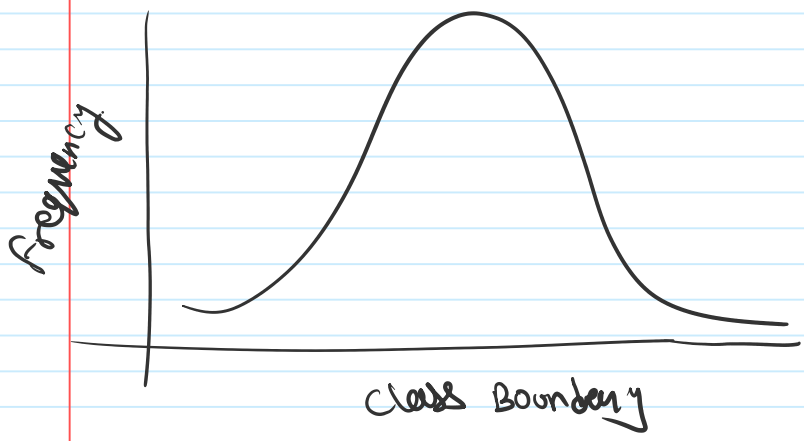
⇓

Smooth curve obtained by joining midpoint of upper side of rectangles in histogram

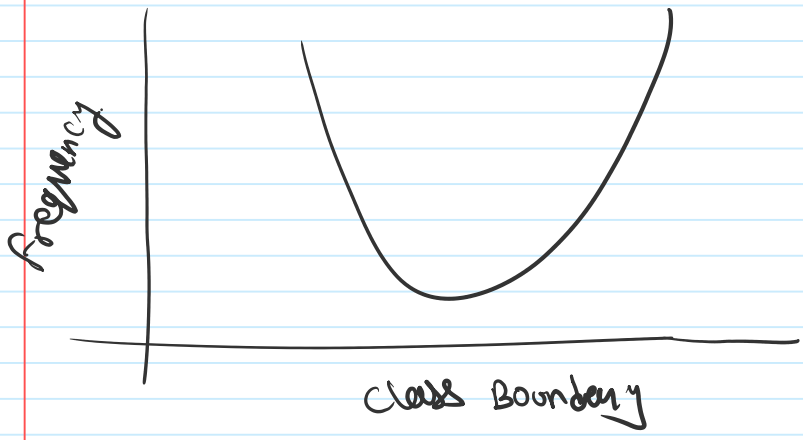
⇓

Area under frequency curve is one unit

Area under frequency curve is one unit

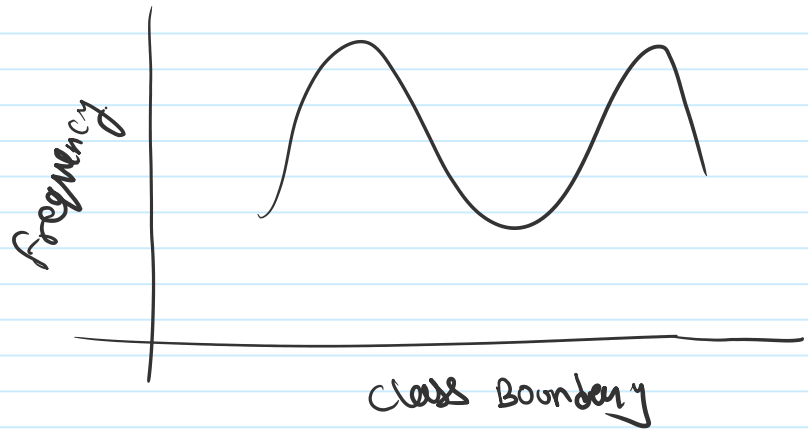


Bell Shape
(weight, profit)



U-Shape
(Passengers in a metro)

J-Shape



Mixed curve