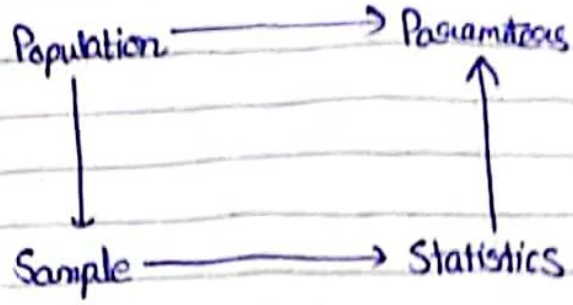


Presentation of Data:

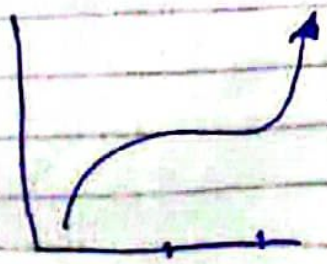
↳ Importance in real life:
- ↳ Restaurant Example
- ↳ Two types of restaurant
- ↳ Preference (not food but representation)

⇒ Job interviews

⇒ FYP Groups
- ↳ Presentation matters more than the content

Presentation of Data

- Tabular
  - *Frequency Distribution
- Graphical
  - * Bar chart
  - * Pie chart
  - * Histogram



Country's progress
↳ layman can understand this

packing outlet.
→ no. of cases in multiple
↳ no. of patients in hospital every day

Data:

Marks
23
25
27
71
78
69
75
32
39
48
⋮

(Data) Age

3
5
7
15
25
19
43
25
24

↳Write a note on Percentile, Quartile and Decile - Assignment #02

① Definition
② Formula
③ Importance
④ Applications

Mode:

Value that is repeated most often times is called mode -

Salesman example.

1, 1, 2, 1, 2, 2, 3, 4, 4, 3, 1, 1

Mode = 1

↳ Processed Data

Mean of Grouped Data is not possible in Real life - (also its approximate value)

↳ Always use raw data/ungroped data for mean or any computation.

In Programming:

Find median of 500 observations -

↳ only sort 1st 250, 251 observations.

others will waste time

500

[250
251

↳ Halt arrangement

↳ If data is Quantitative:

↳ Then use mean, median

↳ If data is Qualitative:

1, 1, 1, 1, 2, 2, 2

Mode = 1 (most saled product)

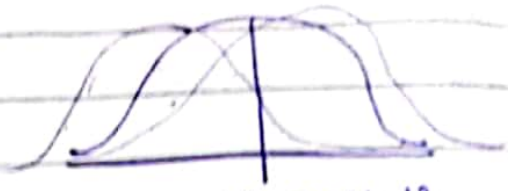
Mean = $\frac{10}{8} = 1.3$

↳ If data is of gender then it is meaningless.

↳ Why mean is more meaningful?

① It is balancing point

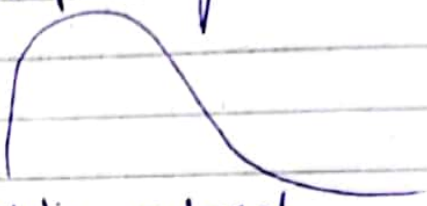
② It is mathematical process that can be processed more.



Mean = Median

↳ Mean is prefered over mean

If shape changed then



↳ Median prefered

↳ If difference between mean & median is less than prefer mean -

30
⋮
40

Mean 35 Median 39

Weighted Arithmetic Mean:

X	W	XW

$\bar{X} = \frac{\sum x}{n} X$

$\bar{X}_w = \frac{\sum W_x}{\sum w}$

↳ When the weights vary per item prefer this

Infinity: (Humary liye 35 rooms khaali infinity hai) Absolute → Nothing

Aesi cheez jo kisi particular condition jahan per relatively mushkil hoo. Impossible relative to that

↳Data ki description ke bagair us ke baray main kuch nhi key skty

Discrete value: (koi cheez exactly define krna)

↳size in bits

↳number of eggs

↳# of stars

↳countable finite

uncountable finite

1,2,3,4,5 — — — — ↳traces in parking lot

students

2
3
4
5
6 → # marks then continuous

Discrete

↳Data ki description ki bagair discrete aur continuous ko hum nhi bta skty.

Continuous Graph: ⇒ Marks

↳size in Bytes

↳height

↳weight

↳speed

0
0.00001
0.0000111
⋮

5

↳Time

↳Qualitative main data string form main

↳Statistics & computer does not deal directly with Qualitative data

① Collection of Data: (Types)

↳According to sources

Primary data: → consumes time & cost

↳First handed Data or data that is collected from first time is called Primary data

↳Through observation, questionaire, etc.

Secondary Data: ↳copied ↳gathered data

Second handed data or process data or data after applying statistical method

↳Through hospital, education institute etc.

↳computer scientists prefer Secondary data.

Code	Gender
M=1	1
F=2	2
	2
	1
	1
	2
	2
	1

Jo likhna katy hain hum

↳computer can't understand strings bcz it works in binary.

↳In order to deal with Qualitative data, we assign numbers to the data. (Qualitative coded data) → numerical data

↳Process data ko reprocess nhi krskty

↳Humesha data raw form main use krna hai bcz tab primary ho ya secondary

↳Statistics does not deal directly with Qualitative data

Qualitative Data (string form) ⟷ Quantitative Data (numerical)

Qualitative:

↳Hum khush hain lekin ye nhi pta k kitny khush hain.

↳Agr 90 level (Quantitative) ko Qualitative main krna to phir levels bana lena

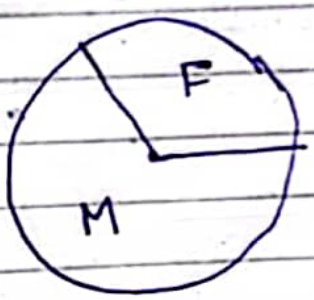
Division wise Data

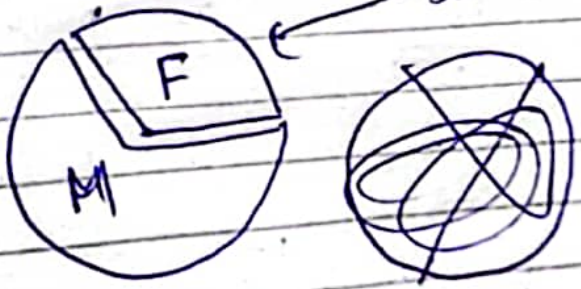
Age	Grade
10–20	A
30–40	B
40–50	C

Pie Chart:

- Alternative of simple Bar Chart

M	10	$\frac{10}{15} \times 360^\circ = 240^\circ$
F	5	$\frac{5}{15} \times 360^\circ = 120^\circ$
	15	





	(X)	(W)	
Subject	CGPA	Credit hrs	xw
S1	2.5	4	10
S2	2.8	3	8.4
S3	3.1	2	6.2
BT	3.9	2	7.8

If weights (changed)

2
3
3
4

CGPA ($\bar{x}_w$) → changes
CGPA ($\bar{x}$) → remains same
$\bar{x}_w = 3.2$ (correct)

$\bar{X} = 3.075$

$\bar{x}_w = \frac{32.4}{11} = 2.945$ (CGPA)

If W [equal weights] then special case

Q_1 Median: 1, 2 → Q_1
Median: 3 → Median = Q_2
4 → Q_3
5

↳ Agr values split na ho rhi hoon to simple median wala concept laggy ga

Apps

	(X)	(W)
App	Ratings	Rated by People
A	4.8	2
B	3.9	5
C	4.2	200

4 { 0.11 ft ... 8 ft

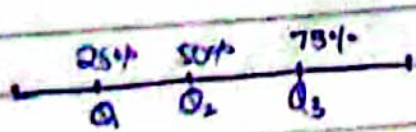
Methods to measure dispersion:

① Range:
Max − Min
5 − 1 = 4 (1, 2, 3, 4, 5)

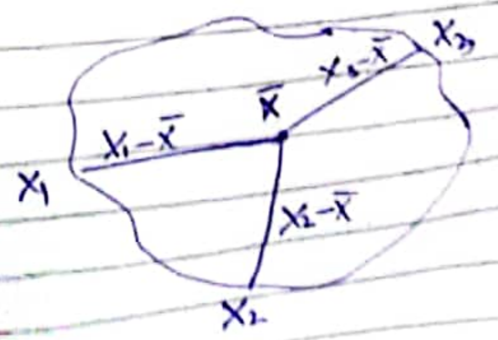
↳ Maximum distance between two points of the data
↳ Minimum distance can't be found.

⇒ **Measure of Dispersion:**
↳ Average sy baaqi values kitna phelao pe hain isko dispersion kehtay. Akela avg sy decision nhi lyy sktay.

9/03/2022



65 score in NTS
92% Percentile
↳ It means 92% students are lower that 65.
↳ Kis level par khary hain ya humain percentile batati hai.



→ X − X̄

Why studying Probability:

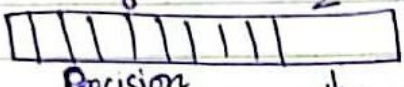
↳ Back-end per chezon ko change krna aur parameters ko samajhna -

14/02/2022

Prediction: (uses Probability)
↳ Measurement of Chances
↳ Prediction of chances (if probability < 50 than NO)

Margin of Error: (Bussiness main loss)
Agr dozen eggs liye aur agr 2 toot bhi jayen to phir ye margin of error

Precision: (Bussiness main Profit) (Accuracy)
Agr 8 eggs sy kam reh gy to beach nhi kr skty to phir ye precision hogi.



↳ bAgr data available na hoo to phir fake data by algorithm k correct honay ka nhi pta chally ga

2 types of statistics:

① Descriptive Statistics
↳ Data Describe in terms of graph or some numerical quantities.

② Inferential Statistics
↳ Procedure

Sampling:
① Blood Sample
② Gori-Surveys for project
③ Hotels feedback

What is data: Raw facts & figures is called data.

① Constant Data: Statistics are not used
↳ no need for any kind of model

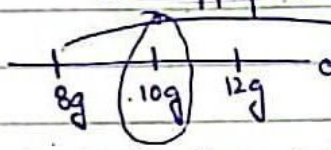
2/2 (1st year) | 2/2 2 (2nd year) | 2/2 2 (n th year)

Sun rises from the east (always same)

② Variable Data: Statistics are used only in this

↳ cut off point

Margin of error = 1
Precision = 2 → 3
→ Vernier Callipers
→ Have quantity main ye dono hoongy



Margin of Error = 2 CGPA.

Population: → Each & Every

Jahan per census krna mushkil ① Time ② Cost

↳ we can't study population in real life so we consider sampling.
↳ Long term main chezain adjusted hoti hain
↳ Life main mushkil bhi aur asani bhi lekin long term main sub control. Bussiness main bhi aesay hy.

Post Analysis: (Postmortem Analysis)
Compare results and probability.
↳ before result

Is sample reliable?
① Take randomly
② Not Biased
↳ Theoretical approach
↳ Not Exact

→ ① Giving correct Data
② How to take data from people

① Data (Variable)

Qualitative
↳ in terms of Quality
↳ gender (male/female)
↳ Hair colour (Blue, Black, Brown)
↳ categorizable
↳ types of car
↳ cities
↳ Grades

2009
↳ 2nd Floor → Room#9
↳ Qualitative

2009
↳ Year
↳ Quantitative

Quantitative: (count)
↳ in terms of numbers
↳ age
↳ salary
↳ Height
↳ Marks
↳ ID card number

↳ Values ko dekh kar nhi bta skty

Quantitative

Discrete
A variable that can take some specific value within an interval
2, 4

Continuous
A variable that can take all values within an interval
1, 2, 3, ...

28/02/2022

	f
CS A	41
CS B	39
CS D	31

M	31
F	10

Aggregate Data

↳ Here we consider (Raw) data → Jisky table ya graph na banaya hoa. Jo process na huwa hoa.

Graphical Ways:

Qualitative Data (not for Quantity) → Bar chart
→ Pie chart

Bar Chart Types

① Single — Multiple — Component

Single variable

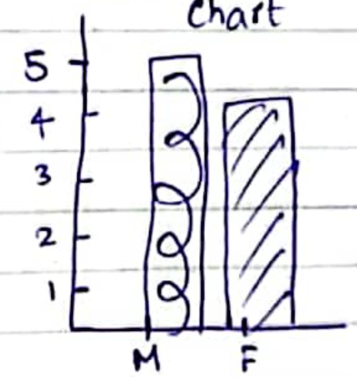
1

Gender
M
M
M
F
F
⋮

Raw Data

	f
M	5
F	4
	9

Single Bar chart

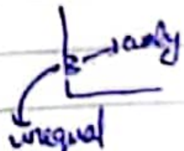


1 variable

CSA	48
CSB	31
CSC	29
CSD	33

(can be multiple)

↳ must be equidistant



↳ laptops companies comparison
↳ Medicine accessory sale of companies

$$\frac{\sum (X-\bar{X})}{n} = 0$$

↳ Har value apny center sy on average kitny faasley pe hai.

1
2
3
4
5

Mean Calculation:

$$\bar{X} = \frac{15}{5} = 3$$

$$\frac{(1-3)(2-3)(3-3)(4-3)(5-2)}{5} = 0$$

Hence it proves, (statistical approach) sum of the deviation of each observation from their mean = 0

$$\sum (X-\bar{X}) = 0$$

$$\frac{\sum |X-\bar{X}|}{n} = \text{Mean Deviation}$$

$$\frac{\sum (X-\bar{X})^2}{n} = \text{Variance}$$

$\sqrt{\text{Variance}} = \text{Standard Deviation (SD)}$

↳ values apni average sy on average kitny faasly per hain

1
2
3
4
5

$\bar{X} = 3$

X	$(X-\bar{X})$
1	4
2	
3	
4	
5	

SD = 1.41

$\bar{X} \pm SD$

3 ± 1.41

1 rg ± 1 g

9 — 11

when SD = 0 & variation = 0

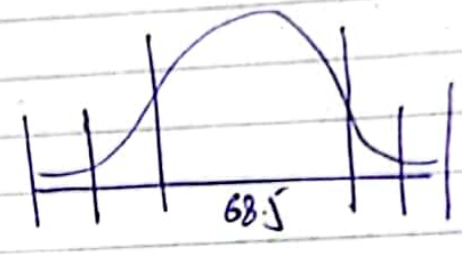
↳ Data is constant

↳ 3 Sigma limits:

68.5% = $\bar{X} \pm 1SD \Rightarrow 3 \pm 1.41$

95.7% = $\bar{X} \pm 2SD \rightarrow 3 \pm 2(1.41)$

99.5% = $\bar{X} \pm 3SD \rightarrow 3 \pm 3(1.41)$



50 ± 5

① 45 — 55

② 40 — 60

③ 35 — 65

50 ~~— 70~~

50 ± 20

① 30 — 70

②

③

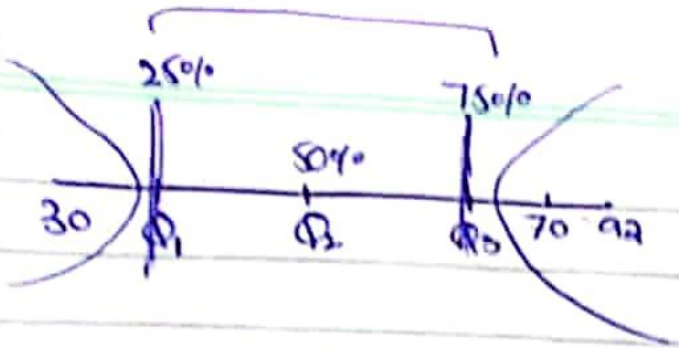
↳ Characteristics & guarantees

↳ Interquartile Range (I.Q.R):

$IQR = Q_3 - Q_1$

70 − 30 = 40

92 − 30 = 62



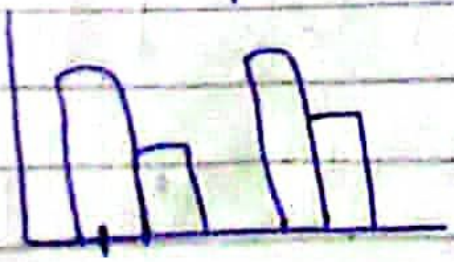
Mean ± S.D

Median ± IQR

Mode ± Range

↳ Qualitative wise Quantitative Analysis

Gendar			Masks
	Avg	SD	
M			
F	—	—	



Assignment # 1

Applications of Statistics in CS

↳ ① Where
② Tool
③ How (purpose)

① ↳ Probability and statistics for Scientific & Engineering (Walpole & Meb...)

② ↳ First course in Probability (Groose & Sluer)

③ ↳

Population: (Generalization) × → **Parameter:** Characteristics of anything is called param...
→ Human characteristics
↳ characteristics of population

For examples, parameters of Board marker is black coloured, black needle, etc.

y = mx + c
fun( ) { }

↳ Parameters are used to enhance and understand things –
↳ Jab tak population available nhi tab tak usky parameters bhi available nhi hongy

Sample: → **Statistic:** → Statistics

Statistic
- Descriptive
  - Graphic + Representation
  - Summary Statistics
- Inferential: → Population → Parameter
  - Sample → Statistics (generalization)

Inferential statistics

Avg CGPA
2.81
Sample Avg

29%

If we have qualitive data,

		City
1	FSD	1-2
2	CHT	2-4
3	MLT	5-6
4	LHR	
5	GSB	

X

then

		frequency of
1	FSD	15
2	CHT	5
3	MLT	3
4	LHR	7
5	GSB	1

M	15
F	5

Gender
M
M
M
F
F
F
F
F

If data is too much for cities then make it categories of homogeneous data (provinces)

Dimension Reduction:
↳ Reduce 50k variables to 50
↳ Machine learning
↳ Big Data Concept

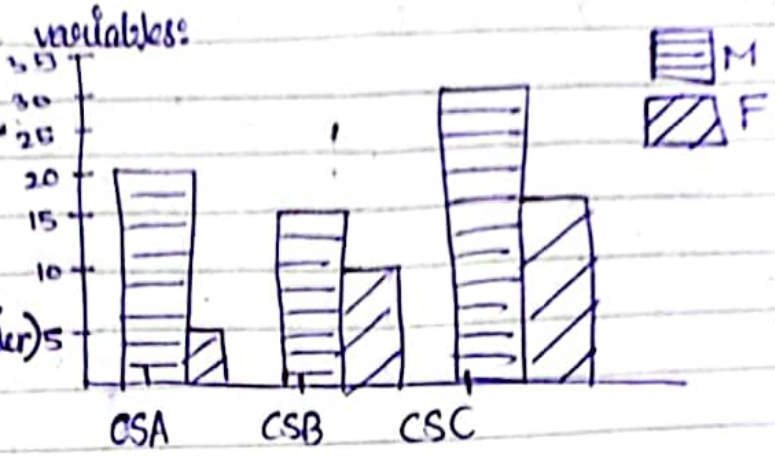
Multiple Bar chart:

↳ only 2 qualitative variables:

	M	F
CSA	20	5
CSB	15	10
CSC	30	15

↳ categories
↳ variable is 1 (gender)

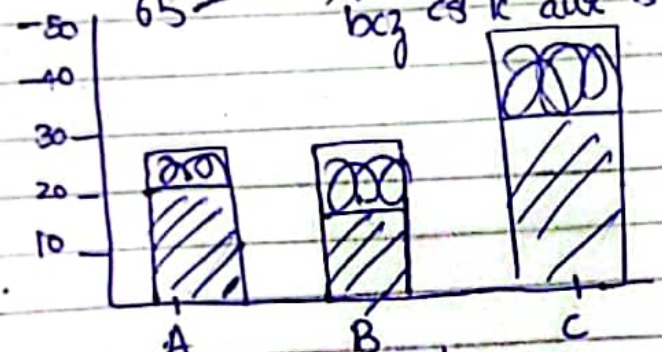
↳ Comparison always demographic



Component Bar Chart:

↳ Quatitative Data
↳ 2 variables

	M	F	Total
CSA	20	5	25 → complete total
CSB	15	10	25
CSC	30	15	45
	65		→ not complete total bcz CS & aur bhi sections



	Pic	Doc	Total → not complete here since we also have texts
Inbox			
SPAM			

↳ Here total is complete (Meaningful)
Multiple bar chart

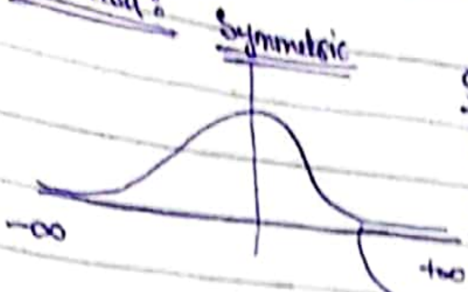
use multiple bar chart ↑ — not meaningful

	GDP	Crime Rate	Total
Pak			
AFG			
IND			

Total → Meaningless here

12/03/2022

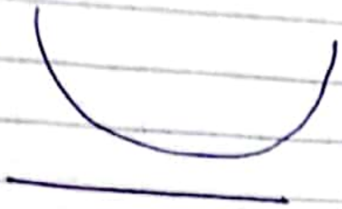
Normal:

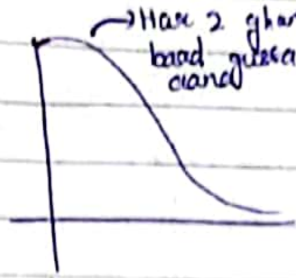


Symmetric curve
Bell Shape curve
Normal curve
→ asymptotic curve:
→ Jo bhi common sy hath rahy chezain hoo thi kam

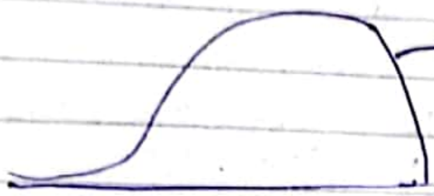
Abnormality:

↳ Bulk of data decides





→ Har 2 ghantay baad gussa aana



→ Har 1 ghanty baad hansna -

↳ Adhy sy ziada ka 4 ⇒ ye abnormal hai

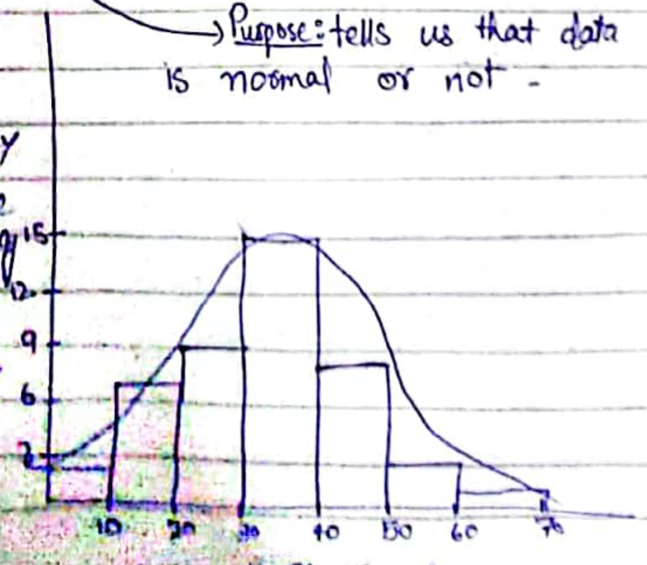
↳ wo chezain jinki existence kam hai yahan lekin wo ziada hooni lag jay to abnormal hain aur highlight hojaye ga -

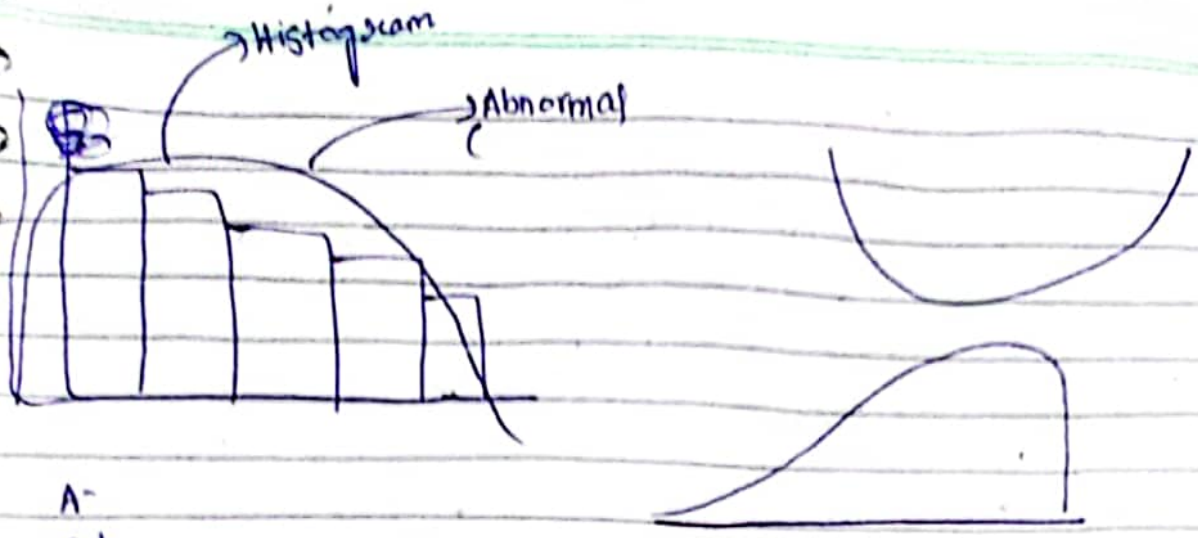
Histogram: ⇒ Quantitative Data (Presentation of Data)

Histogram → Tells us about the shape of the data
→ Purpose: tells us that data is normal or not -

Age	f
0-10	2
10-20	7
20-30	9
30-40	15
40-50	8
50-60	3
60-70	1

frequency or Relative frequency





A⁻
A⁺
AB
B⁺
O⁻

3 types of Average:
① Mean ② Median ③ Mode

Asithmetic Means (AM)
↳ Mean ↳ Average
↳ Representation: $\bar{X}$

Median:
↳ Middle value of arranged data

1
2
3 → Median
4
5

(1, 2, 3, 4 → Median = $\frac{2+3}{2} = 2.5$

↳ Not effective by extreme observations
↳ Not based on all observation
↳ Make 50/50 Division

$$\bar{X} = \frac{\text{Sum of all observations}}{\text{Number of observations}}$$

Σ = Summation - Sum of all
Π = Capital Pie = Product of all

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

↳ based on all observations

1
2
3
4
5

$\bar{x} = \frac{15}{3} = 5$ ↳ Average is affected by extreme observations ⟹ Median = 3

$\bar{X} = \frac{60}{3} = 12$ ⟹ Median = 3

$\bar{X} = 120$ ⟹ Median = 3

↳ Balancing point

Frequency distribution:

Main Aspects:
- Size of class (no restriction) ⇒ just suppression
- No. of classes (restriction) ⇒ (6 or 7)

Classes	
0–10	≤0 <10
10–20	≤10 <20
20–30	
30–40	
40–50	
50–60	
60–above	

0–50	150
50–100	50

] → merging

1–2	2
2–3	2
4–5	

] → scattered

V.V.Imp
↳ must be 6 or 7 total classes

$$\frac{\text{max} - \text{min}}{6} \Rightarrow$$ ~~number of data~~

⇒ Always Equal Data

If classes
0–10
11–20
21–30
then continuous graph becomes discontinuous

classes	Frequency	Commulative Frequency (ΣF)	Relative Frequency
0–10	2	2	2/45 = 0.044
10–20	7	9	7/45 = 0.155
20–30	9	18	9/45
30–40	15	33	15/45
40–50	8	41	8/45
50–60	3	44	3/45
60 & abave	1	45	1/45
	45 observations		

45
Distributing data Accord to their frequency.
43
36
27
12
4
1

Age Below 50 — Inverse Commulative Frequency — Age Above 50

Multiply it with 100 we will get percentage