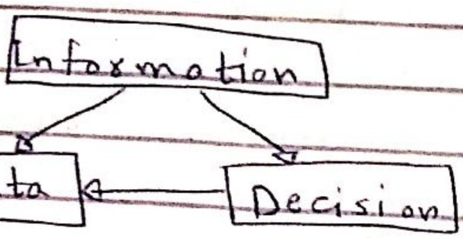
Probability & Statistics - MT2005

Lec#01

Statistics:-

- Collection of data.
- Presentation of data.
- Analysis of data.
- Decision about data.



Lec#02

Data:-

Raw facts and figures are called "data".

Population:-

Population is the entire set of items from which you draw data for statistical study. It can be a group of individuals, set of items.

Noun as census, enumeration,

Cost ↑
Time ↑

Common person who did not Knows about statistics	A person who Knows about statistics
Collection of data not well.	Collection.
Presensation.	Presentation.
Analysis of data not well.	Analysis.
Decision.	Decision.

Lec# 03

Sample:-

Sample is the part/chunK/sub-part of population. Which contains the whole characteristics of the population.

- Sample based studies are called sample-surveys. e.g., In CS how much students have glasses, in class of 39,

No. of students with classes = 19

Total students in class = 39

Result $= \frac{19}{39} = 49\%$

Data:-

Raw facts & figures are called data.

Types of data:-

1) Qualitative:-

Data about quality or characteristic of variable. e.g., name, color, etc.

2) Quantitative:-

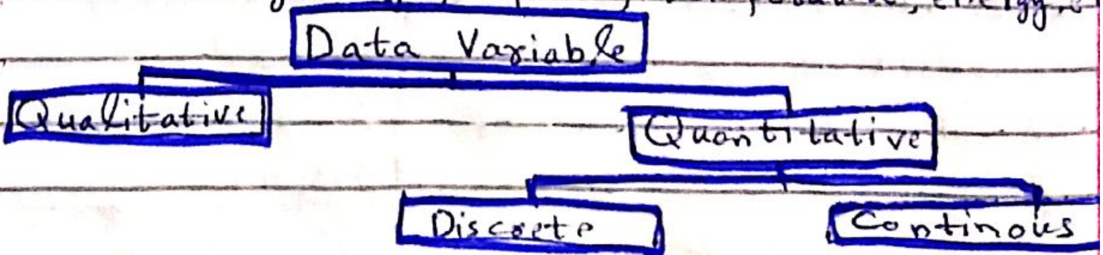
Data about numeric variable or can be represented in numeric form. e.g. height, weight, CGPA, etc.

2.1) Discrete:-

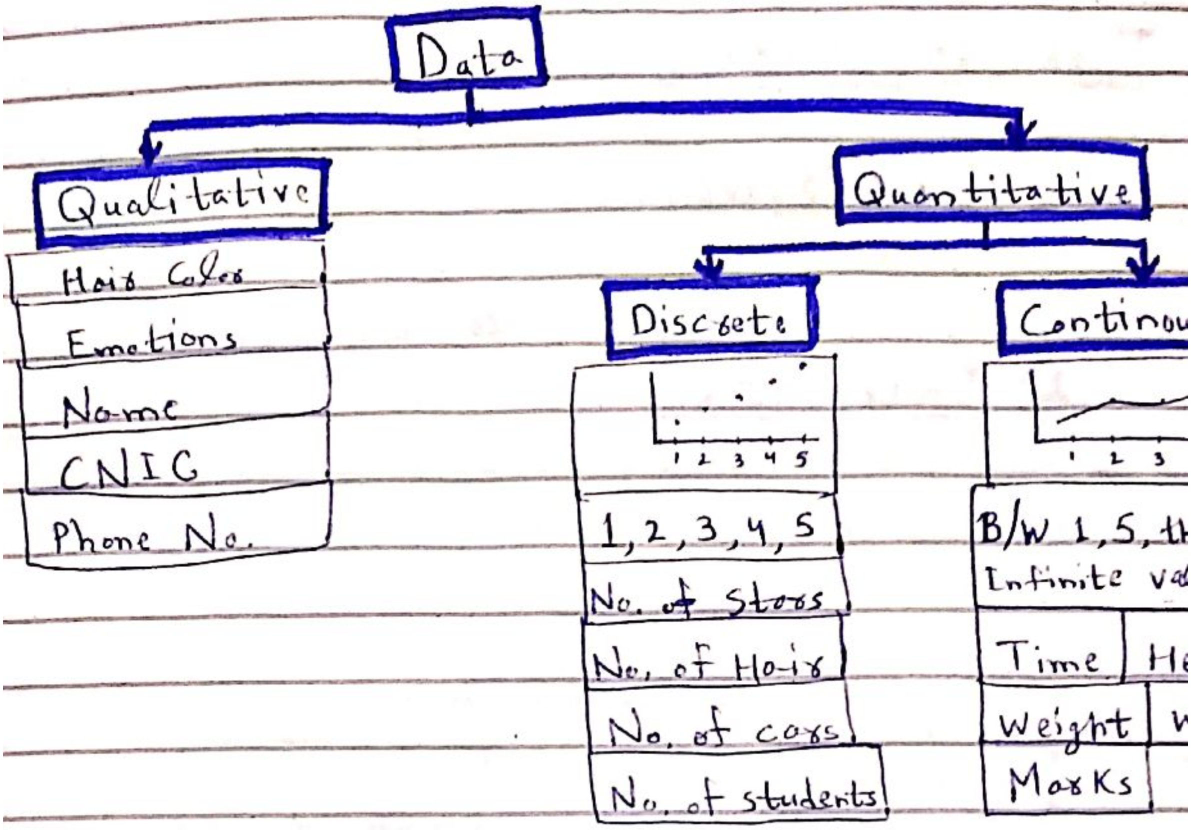
A variable or numeric variable that is obtained by counting. e.g., no. of present students.

2.2) Continous:-

A variable that is obtained by measuring. e.g., speed, temperature, energy, et



Lec#04



- We cannot perform statistics method to qualitative data. So, we assign code to them. e.g.,

Gender	Gender.
M	1
F	2
M	1

Assign,

M = 1

F = 2

- We can convert quantitative data to qualitative like, marks to grade, IQ Level
- We cannot convert qualitative data to quantitative data. e.g., grade to marks is not possible.

Lec#05

Types of Statistics:-

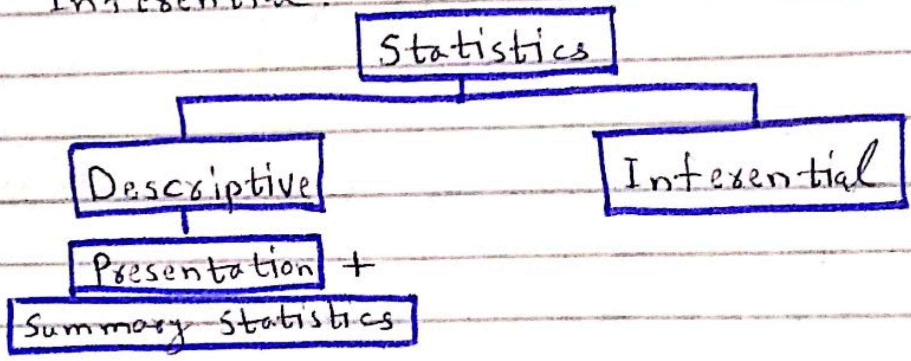
1) Descriptive:-

How to describe/Present data/method

- Presentation + Summary statistics.

2) Inferential:-

Generalization of sample statistic towards population parameter is called Inferential.



Parameters:-

The characteristics of anything is called parameter (from population).

Statistic:-

The characteristics of sample data.

Types of Collected Data:-

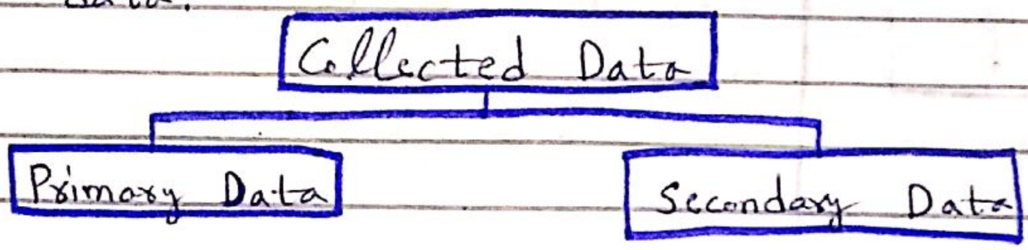
1) Primary Data:-

- The data that is collected first time, or first handed.
- Raw form of data.
- More time consuming, but with your own choice.
- Collect data by yourself, don't

- Usually, questionnaire used to collect data.

2) Secondary data:-

- When any statistical tool is apply on primary data then it becomes secondary data.
- Second-handed data.
- Less time consuming, but we don't have choice.
- Ready-to-use data, and we have to compromise on it.
- Use websites to collect organize data.



Lec#06

Presentation of Data:-

Presentation of data has a very big role in our daily life.

Types of Presentation of data:-

1) Tabular.
2) Graphical.

1) Tabular:-

- Frequency Distribution:-

Note: It is all for Quantitative data means classes.

- Classes:-

1. All of data should be there in the range of classes.

- Two Important things about classes are:-

1) No. of Classes : 6 or 7

2) Size of Class : Equal Gap

- Equal Gap should be there in the size of classes to make algorithm, etc because different gap is subjective thing & can't make algorithm, etc.
- To check/to know the equal gap for given data:-

$$\frac{\text{max} - \text{min}}{\text{No. of Classes}} \quad \text{e.g.,} \quad \frac{\text{max} - \text{min}}{6 \text{ or } 7}$$

Wrong Approach

Classes
0 – 50
50 – 100

Classes
0 – 2
2 – 4
4 – 6
6 – 8

Wrong Approach

Classes
0 — 10
11 — 20
21 — 30

Classes
0 — 10
10 — 20
20 — 30

Less than | <

Greater than Equal to | > =

OR Vice Versa

Example:-

Age Counts/No.

Classes	Frequency (f)	Commulative Frequency (c.f) Normal	Commulative Frequency (c.f) Inverse	Relative Frequency (r.f)
0 — 10	1	1	30	1/30 = 0.033 = 3.3%
10 — 20	3	4	29	3/30 = 0.1 = 10%
20 — 30	5	9	26	5/30 = 0.166 = 16.6%
30 — 40	10	19	21	10/30 = 0.333 = 33.3%
40 — 50	6	25	11	6/30 = 0.2 = 20%
50 — 60	4	29	5	4/30 = 0.13 = 13%
60 above	1	30	1	1/30 = 0.033 = 3.3%

Gender	f
M	10
F	3

- Qualitative data don't have classes. We cannot assign codes to qualitative but we can do all others. Grouping of cities into divisions or provinces,

2) Graphical:-

Considered the type of data while collecting.

Methods:-

1) Bar Chart:-
   1. Simple bar chart.
   2. Multiple bar chart.
   3. Component bar chart.
2) Pie Chart.
3) Histogram.

1) Bar Chart:-

• Bar Chart are just for qualitative data.

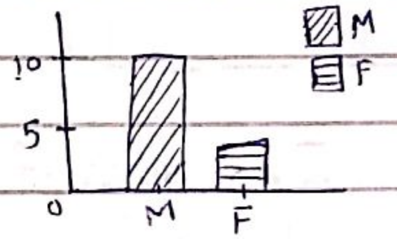
1. Simple bar Chart:-

• Only for one variable of data.

①

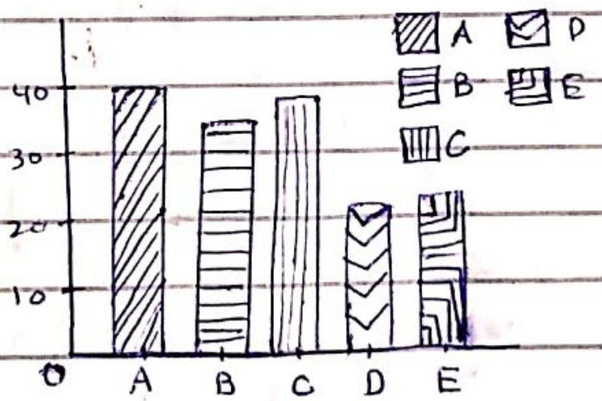
Gender	f
M	10
F	3

→



②

Section	f
A	40
B	35
C	38
D	21
E	23

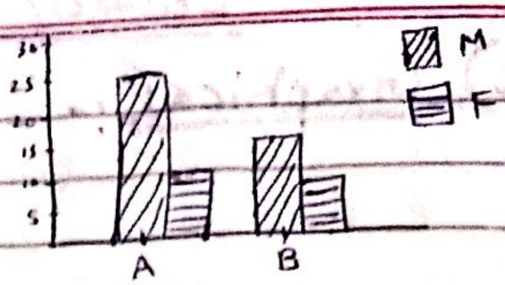


2. Multiple bar Chart:-

• Used for two variables of qualitative data e.g., section with gende

→

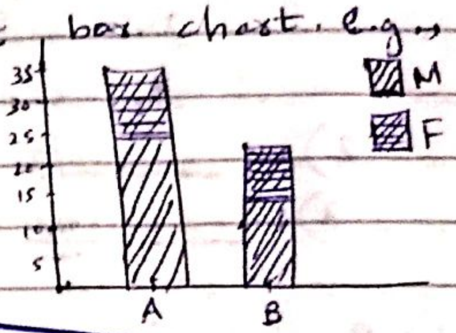
Section SE	M	F
A	25	10
B	15	07



3. Component bar Chart:-

· Initially same as multiple bar chart. If the total is meaningfull th moves to component bar chart. e.g.,

Section	M	F	Total
A	25	10	35
B	15	07	22



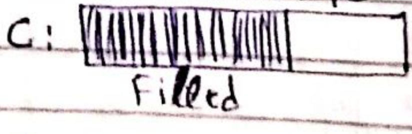
Wrong Approach

	Pak	India	Ban
GDP	—	—	—
NI	—	—	—
AP	—	—	—
Total	—	—	—

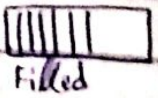
Meaning less Total

Examples of Components:-

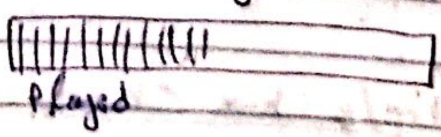
① Hard disk in PC



② Battery



③ Video Player



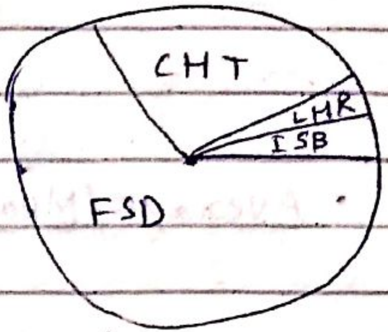
Lec#08

2) Pie Chart:-

Pie chart is only for one variable of qualitative data, e.g., Cities.

Cities	f	Angle
FSD	15	216°
CHT	7	100.8
LHR	2	28.8
ISB	1	14.4
Total	25	

Angle → $\boxed{\frac{\text{City}}{\text{Sum}(f)} \times 360°}$

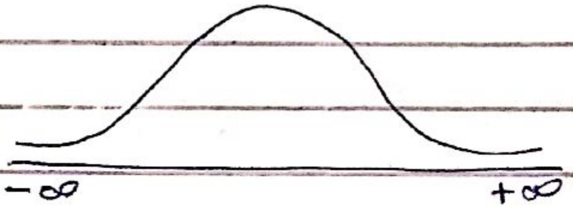


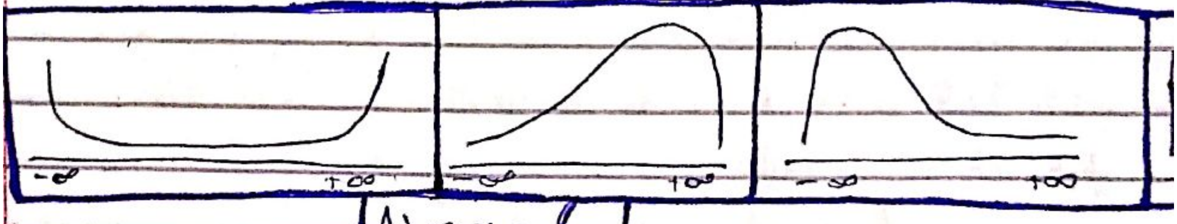
3) Histogram:-

Histogram is for quantitative data.

- Normal:-

Bell shape curve,
Symentic curve,
Normal curve



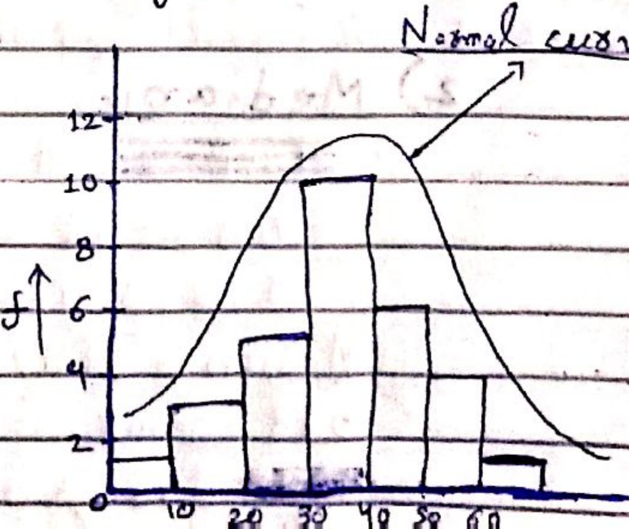


Abnormal

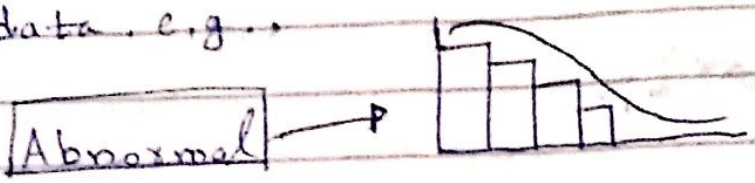
Example of Histogram, e.g.,

classes.	f
0 — 10	1
10 — 20	3
20 — 30	5
30 — 40	10
40 — 50	6
50 - 60	4
60 above	2

→



- Also, Histogram tells the shape of data. e.g..



~~Descriptive statistics:-~~

Summary Statistics:-

- How to summarize the quantitative data. A single value that represents the whole data. Average/Mean.

- **Average/Mean:-**
  - Majority of data.
  - Centered Value.
  - Balancing Point.

- **Measure of central Location:-**

1) **Mean:-**

- Balancing Point of Data.

$$\text{Mean} = \bar{X} = \frac{\text{Sum of all Observations}}{\text{No. of Observations}} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

$$\bar{X} = \frac{\sum_{i=r}^{r} X_i}{n}$$

- But, there is effects on data/Mean by extreme observations, e.g., 1, 2, 3, 4, 500

$\bar{X} = 510/5 = 102$

- Can 102 be a representator of 1, 2, 3, 4, 500. Not/No. Now, moves to median.

2) **Median:-**

- [redacted] Middle most observation.
- Even Count of Data,

$1, 2, \underbrace{3, 4}, 5, 6$

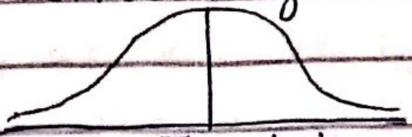
Median = 3+4/2 = 3.5

- Odd Count of Data,

[redacted] 1, 2, 3, 4, 5

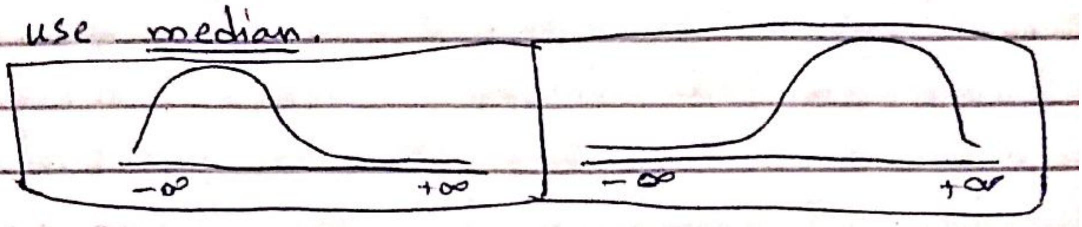
Median = 3

- Median is less sensitive method that connot effect by outlyers.



Still in Normal → $\bar{X}$ = Median

- In case of no outlyer, prefer to use mean, because it has mathemati formula. • But, in given below case use median.



1, 2, 3, 4, 500, then use median.

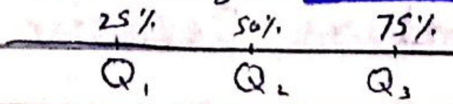
Lec#09

3) Mode:-

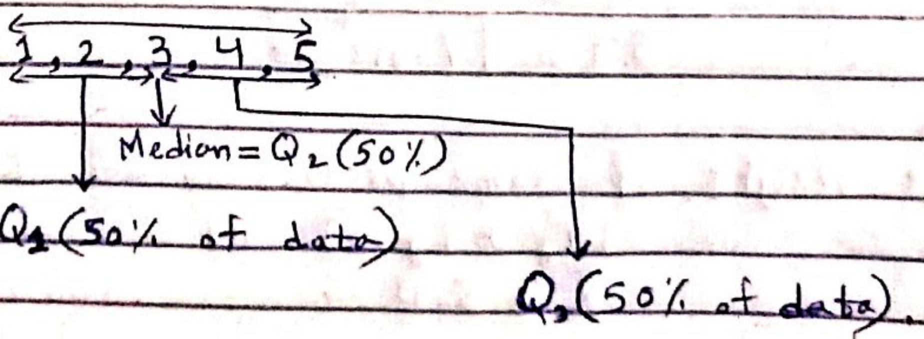
- The most common value in qualitative data. OR
- Qualitative data with highest frequency, e.g., 1, 2, 3, 2, 2, 1, 1, 2, 2, 3, 2

Mode = 2

- Mean, Median, is only 50% of data
- To observe the qualitative or quantitativ data in chunks, e.g., Quantile is used.



, e.g.,



- Average is not enough to analyze the data. To Know the Difference of the data from average. We used dispersion or variation.

Measure of Dispersion :-

1) Range :-

$R = Max - Min$, e.g., 1, 2, 3, 4, 5

$R = 5 - 1 = 4$, this result tells the max distance or difference of data th can occurs.

$R = (x - \bar{x})$

$$\rightarrow \boxed{R = \frac{\Sigma (x - \bar{x})}{n} = \frac{0}{n} = 0}$$

Data	$x - \bar{x}$
1	$1-3=-2$
2	$2-3=-1$
3	$3-3=0$
4	$4-3=1$
5	$5-3=2$

So, Sum of the deviation of each observation from their mean = 0

$$\frac{\Sigma (x - \bar{x})}{n} = 0$$

To overcome this problem, we have :-

1) Variance :-

$$V = \frac{\Sigma (x - \bar{x})^2}{n}$$

2) Standard Deviation :-

$$S.D = \sqrt{Variance}$$

- Weight of same product mentioned as :-

$10g \pm 1g \; (9 - 11)$

$\bar{x} \pm S.D$

		Example
	3 Sigma	10g + 1g
$\bar{X} \pm S.D * 1$	68.5%	9 − 11
$\bar{X} \pm S.D * 2$	95.7%	8 − 12
$\bar{X} \pm S.D * 3$	99.5%	7 − 13

Another Example:- (57 ± 12) Marks

57 ± 12 (12*1)	45 − 69
57 ± 24 (12*2)	33 − 81
57 ± 36 (12*3)	21 − 93

3) Inter Quantile Range (IQR):-

$$\mathrm{IQR} = \overset{\text{Upper Quantile}}{Q_3} - \underset{}{Q_1} \xrightarrow{} \text{Lower Quantile}$$

1) Good Method for quantitative data with no outlier:-

$\bar{X} \pm S.D$

2) Good Method for quantitative data with outliers:-

Median ± IQR

3) Good Method for qualitative data:-

Mode ± (Max − Min)

Qualitative wise Quantitative Analysis

Gender	Age
M	—
F	—
M	—
M	—
F	—
F	—

→

Gender	Age	S.D	—
M	—	—	—
F	—	—	—

Now, we can make graph o
that.