

Introduction to Statistics

Definition :- Statistics is a mathematical science including methods of collecting, organizing and analysing data in such a way that meaningful conclusions can be drawn from them.

Data : facts or pieces of information that can be stored and measured.

Example:- Scores made by Virat Kohli in last five matches
[102, 33, 0, 45, 46]

Example - Heights of Students in the classrooms
[110cm, 150cm, 170 cm - - -]

Motivation | Use of Statistics

1. Weather forecast
2. Sports Analytics - [102, 33, 0, 45, 46] $\rightarrow \min = 0$
 $\max = 102$
 $t_{Avg} =$
3. Election Campaign
4. FMCG | e-commerce
5. Medical | genetics

Types of statistics

① Descriptive Statistics



Describe

Defn → It consists of organizing and summarizing the data | population.

Virat Kohli - [45, 62, 65, 81 - - -]

strike rate

② Inferential Statistics

Defn

Inference.

It consists of using data that has been measured to form conclusion about population.

→ With the given sample data can we conclude something about the population.



① Descriptive Statistics

Example-1 - Virat Kohli Scores

[150, 160, 80, 80, 70, 180, 91, ...]

Avg score / strike rate

Example-2 Avg Height/ weight of students in class.

[150 cm, 130 cm, 130 cm, ...]

Q. What is Avg height of students in class? $\rightarrow \frac{150 + 130 + \dots}{50} \rightarrow$ some avg value.

Q. What is the most frequent height of students in class? $\rightarrow 130 \text{ cm}$.

Q. What is avg score scored by VK in his career?

$$\frac{45 + 60 + \dots}{n} = \text{avg score}$$

Techniques of Descriptive statistics

① Measures of Central tendency (mean, median, mode)

② Measures of Symmetry (skewness, kurtosis)

③ Measures of dispersion (standard deviation, Variance)

② Inferential statistics

Why? \rightarrow Population is large

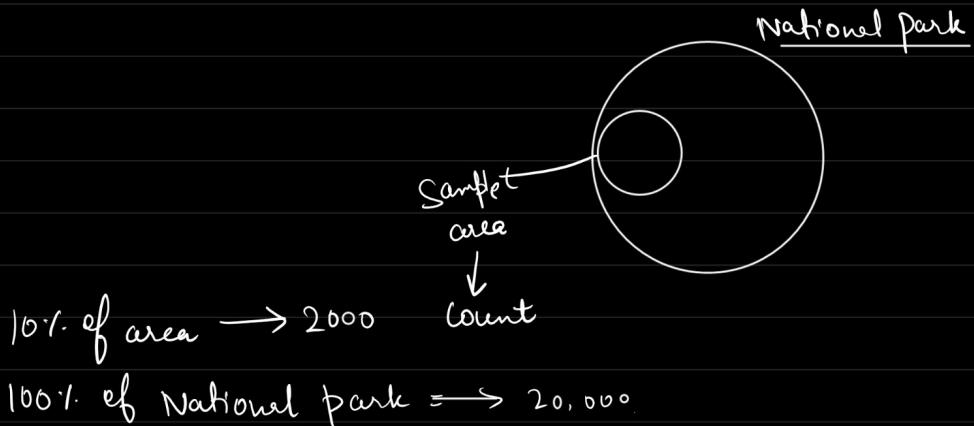
\rightarrow time and resource constraints

Example-1 Avg age of Population of India.



Avg

Example-2 → No of trees in a National Park



Descriptive Question

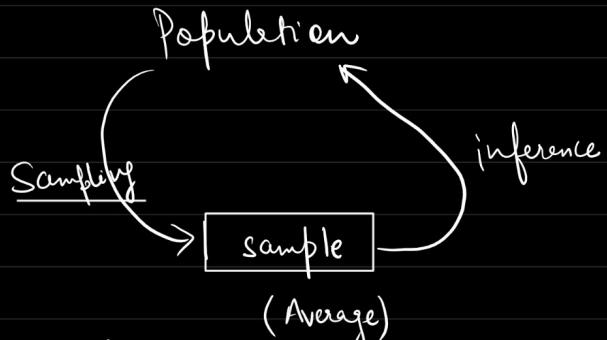
* Classroom

→ What is Avg age/height of students in class?

Inferential Question

→ Are the avg height of students in the class is what you expect in the entire school?

Types of Sampling



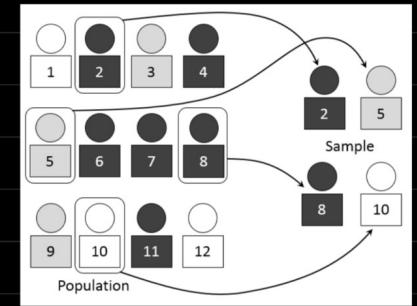
Ex. income of all households in India.
↓

Sampling technique

- ① Simple Random Sampling
- ② Stratified Sampling
- ③ Cluster Sampling
- ④ Systematic Sampling

① Simple Random Sampling

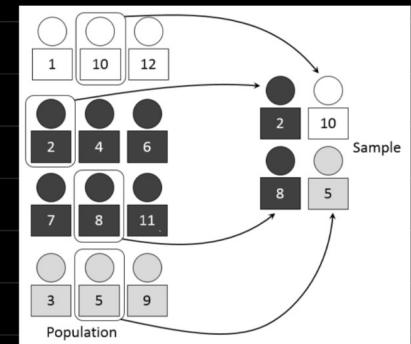
- Every member of the population (N) has an equal chance of being selected for your sample.
- Each person has $1/12$ of being selected in the sample.



② Stratified Sampling

↓
Strata → Layers/groups

Avg height of Population of India:- In SRS the people of North East India might have lesser avg height as compared to rest of India → Stratified Sampling should be used.

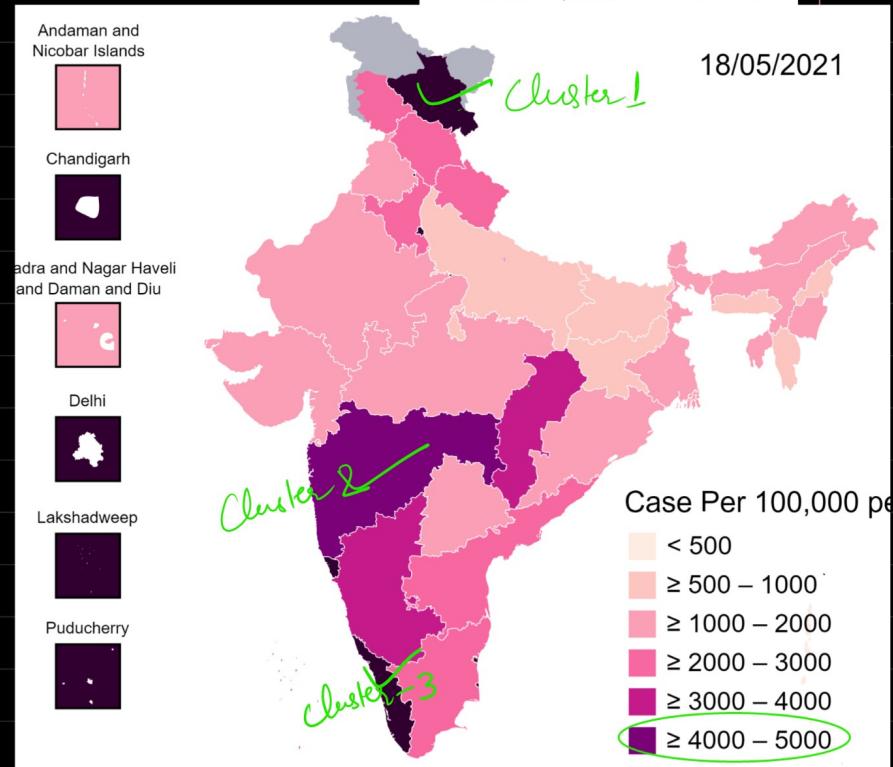
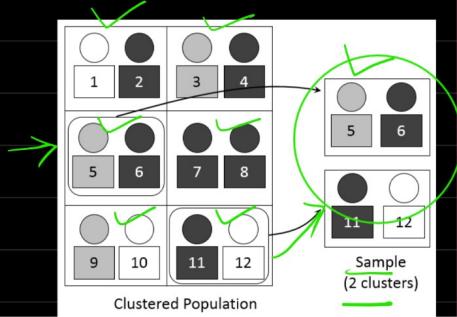


- different distinct categories are there.
- A simple sample would be chosen from each strata or layer.

③ Cluster Sampling

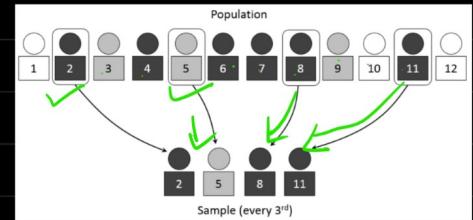
→ divides the population into groups or clusters. Some of these clusters are randomly selected.

→ Then all individual in the chosen cluster are selected in the sample.



④ Systematic Sampling

→ Every n^{th} element will be selected.

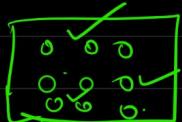


Example → Odd roll no

→ People born on odd years.

Recap

→ SRS



- Quota Sampling
- Min-max Sampling
- Convenience Sampling
- Accidental Sampling

→ Stratified Sampling

e.g. Students of a school — Class 1, 2, 3, ..., 10

e.g. Population of India

e.g. Gender — Male / Female



→ Cluster Sampling

All the data



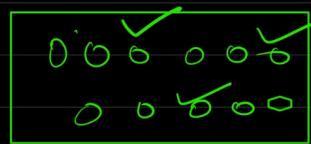
of randomly selected clusters as sample.

→ Systematic Sampling

every n^{th} member

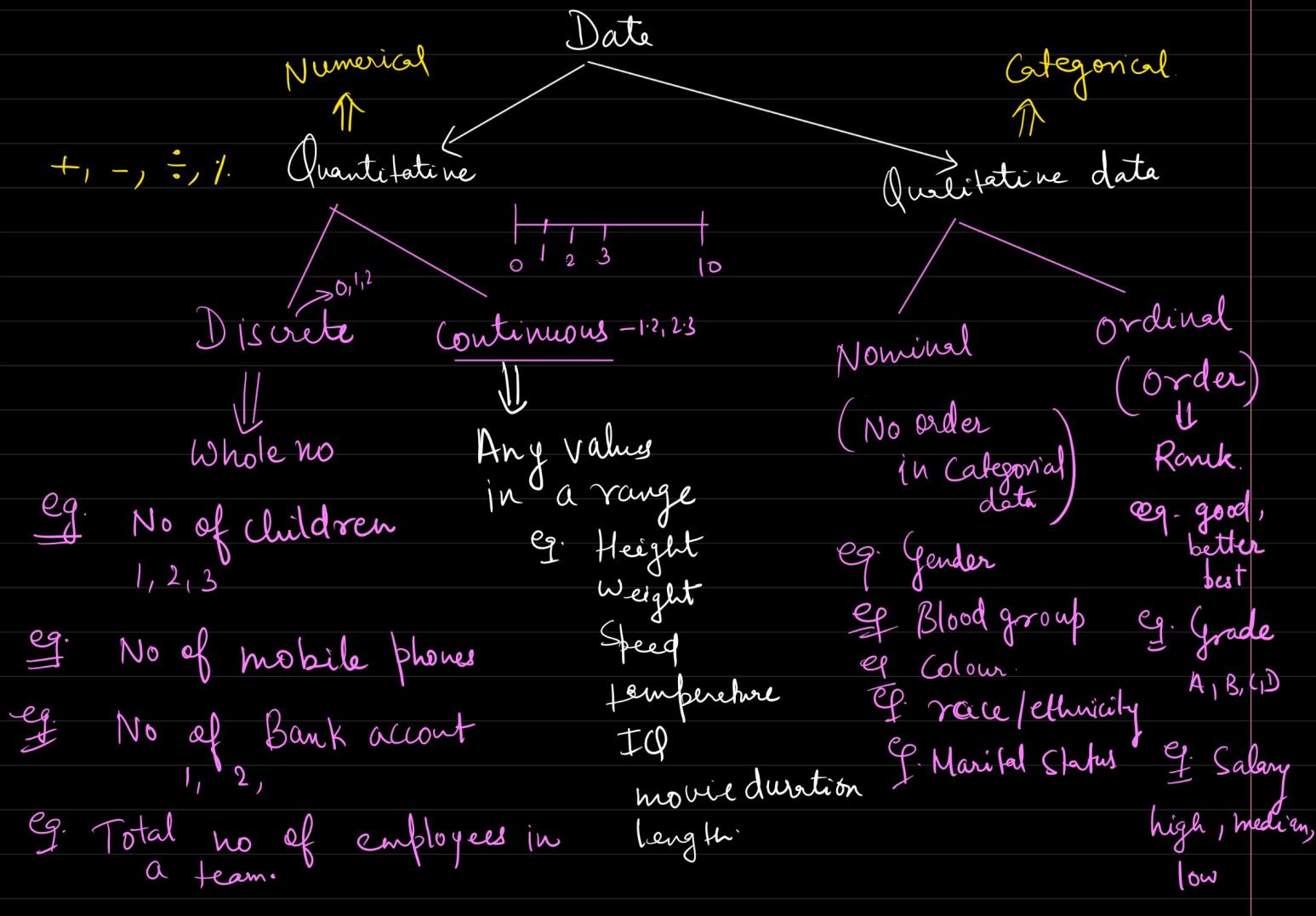
e.g. Students odd roll no

e.g. people born in leap years



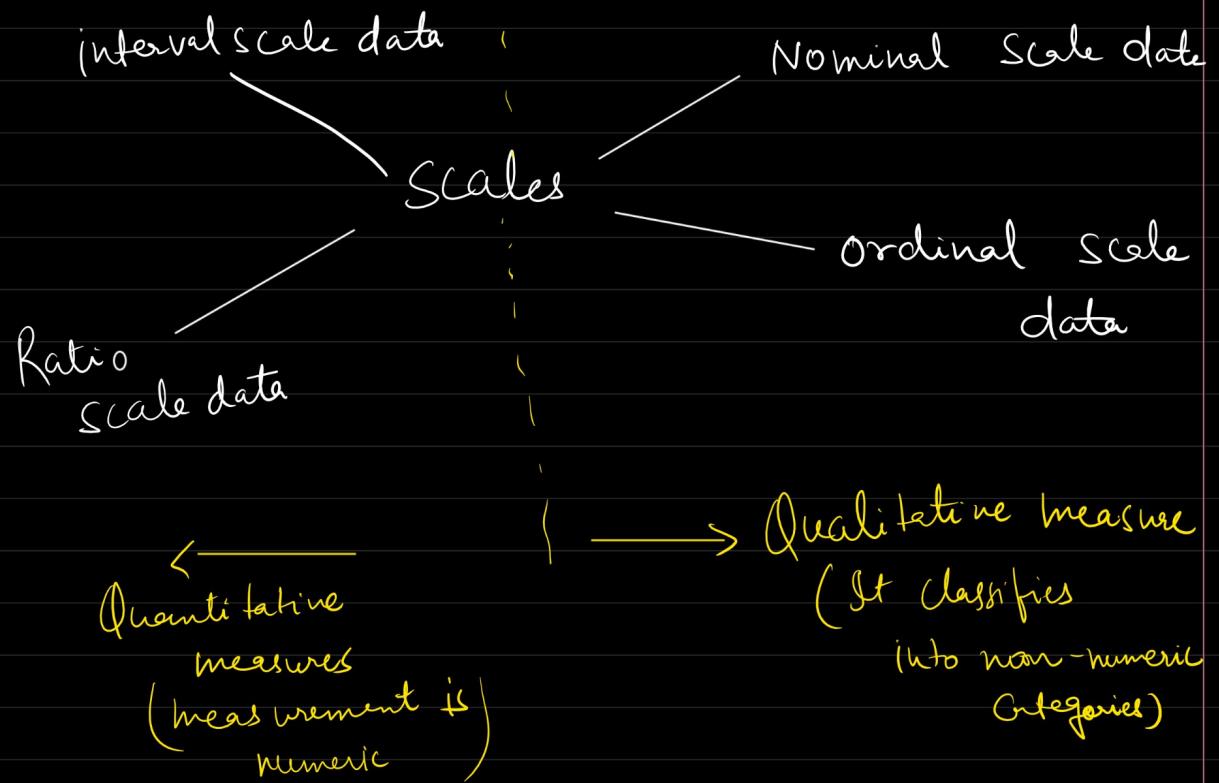
Types of data

Age	TypeofContact	CityTier	DurationOfPitch	Occupation	Gender	NumberofPersonVisi	NumberofFollowups	ProductPitched	PreferredPropertySta	MaritalStatus
41	Self Enquiry	3	6	Salaried	Female	3	3	Deluxe	3	Single
49	Company Invited	1	14	Salaried	Male	3	4	Deluxe	4	Divorced
37	Self Enquiry	1	8	Free Lancer	Male	3	4	Basic	3	Single
33	Company Invited	1	9	Salaried	Female	2	3	Basic	3	Divorced
	Self Enquiry	1	8	Small Business	Male	2	3	Basic	4	Divorced
32	Company Invited	1	8	Salaried	Male	3	3	Basic	3	Single
59	Self Enquiry	1	9	Small Business	Female	2	2	Basic	5	Divorced
30	Self Enquiry	1	30	Salaried	Male	3	3	Basic	3	Married
38	Company Invited	1	29	Salaried	Male	2	4	Standard	3	Unmarried
36	Self Enquiry	1	33	Small Business	Male	3	3	Deluxe	3	Divorced
35	Self Enquiry	1	22	Small Business	Male	2	2	Basic	4	Divorced
	Self Enquiry	1	21	Salaried	Female	2	4	Deluxe	3	Single
31	Self Enquiry	1	32	Salaried	Male	2	3	Basic	3	Married
34	Self Enquiry	1	25	Small Business	Male	3	3	Basic	3	Married
28	Self Enquiry	1	30	Salaried	Male	2	4	Basic	3	Single
29	Self Enquiry	1	27	Salaried	Female	2	2	Standard	5	Married
32	Self Enquiry	1	11	Salaried	Male	3	2	Basic	4	Married
22	Company Invited	1	22	Small Business	Male	3	2	Basic	3	Married
53	Self Enquiry	3	8	Salaried	Female	3	4	Super Deluxe	3	Divorced
	Self Enquiry	1	8	Salaried	Male	2	3	Basic	3	Single
Company Invited	1	17	Salaried	Female	3	2	Deluxe	3	Married	
Self Enquiry	3	15	Salaried	Male	2	4	Deluxe	5	Single	
34	Self Enquiry	1	13	Salaried	Fe Male	2	3	Standard	4	Unmarried
21	Self Enquiry	1	21	Salaried	Male	3	3	Basic	3	Single
34	Self Enquiry	1	12	Small Business	Male	2	3	Basic	5	Single
53	Self Enquiry	1	11	Salaried	Female	2	3	King	3	Married



→ It will take
specific finite
whole no. values.

Scales of measurement



① Nominal Scale data

Example:-

Employees

M

F

M

F

M

M

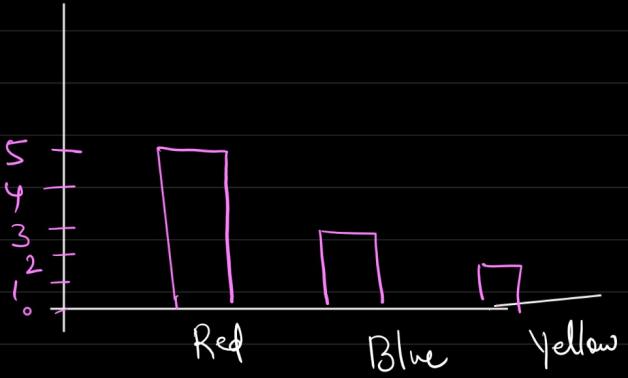
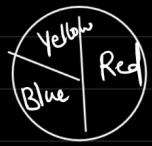
M

→ No Order in the data

$$\begin{array}{l} M - 5 \\ F - 2 \end{array}$$

Example Red - 5 — 50%.

Blue - 3 — 30%
Yellow - 2 — 20%



② Ordinal Scaled data

→ Order and rank matter.

→ Difference can not be measured

Performance
of
students
1st
2nd
3rd

Rank

X [1st
X [2nd
X [3rd
X [4th

ex - Salary
(high, medium, low) 3rd

ex - Educational
qualification 4th rank

1st — 2nd

③ Interval Scale data

→ The rank and order has a meaning

→ Difference can be measured (excluding ratio)

→ It doesn't have 0 starting value

Example → Score 90 — 5
85 — 5
70 — 20
50 — 20

example - Length
Weight

Temperature

inside - 30°C — 0°C
outside - 60°F — -10°C

$$\frac{30^{\circ}\text{C}}{60^{\circ}\text{C}} = 1:2$$

height
temperature
Revenue of a company

④ Ratio Scaled data

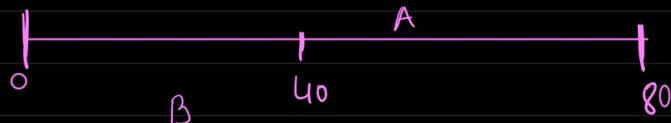
- Order and rank has a meaning
- Differences and ratio are measurable
- It does have a 0 starting point

Ex:

$$\begin{array}{l} A \rightarrow 80 \text{ kg} \\ B - 40 \text{ kg} \end{array}$$

Person A is twice the weight of B

$$\frac{80}{40} = 2:1$$



→ 0 as
true

Starting
point



e.g
height
Weight
Time
Age

Scale of Measurement

Data	Nominal	ordinal	Interval	Ratio
Labelled	✓	✓	✓	✓
Meaningful order	✗	✓	✓	✓
Measurable difference	✗	✗	✓	✓
True zero starting point	✗	✗	✗	✓
Example	Gender Religion Post office code Location	Satisfaction Rating Grade Rank	IQ Temp Score ht wt	Height wt time Age