

Introduction to statistics

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

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

exmp1: Scores made by Virat Kohli in last five matches
[102, 33, 0, 45, 46]

Exmp2 - Heights of students in the classrooms
[110cm, 150cm, 120cm - ...]

Motivation / Uses of statistics

1. Weather forecast
2. Sports Analysis - [102, 33, 0, 45, 46] — $\begin{matrix} \text{min} - 0 \\ \text{max} - 102 \\ \text{avg} - \end{matrix}$
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.

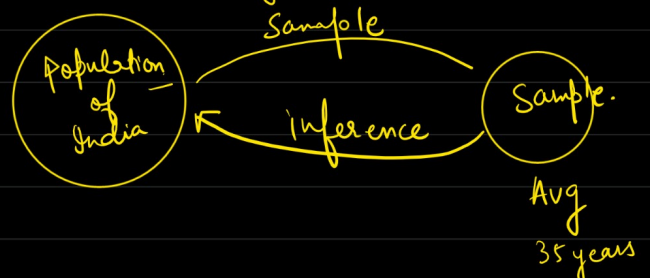
Virat Kohli - [45, 62, 65, 81 - ...]
Strike rate

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

[45, 60, 80, 80, 70, 80, 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? $\rightarrow 130$ cm.

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

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

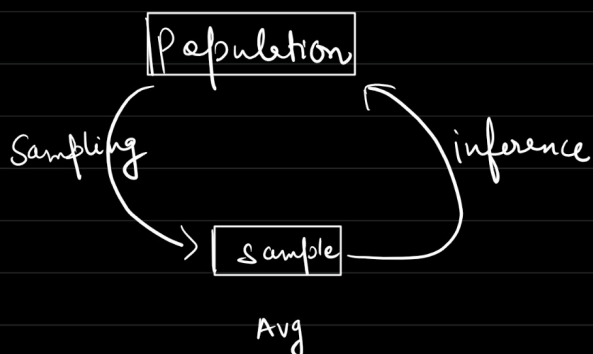
Techniques of Descriptive Statistics

- ① Measures of Central tendency (mean, median, mode)
- ② Measures of Symmetry (skewness, kurtosis)
- ③ Measures of dispersion (standard deviation, variance)

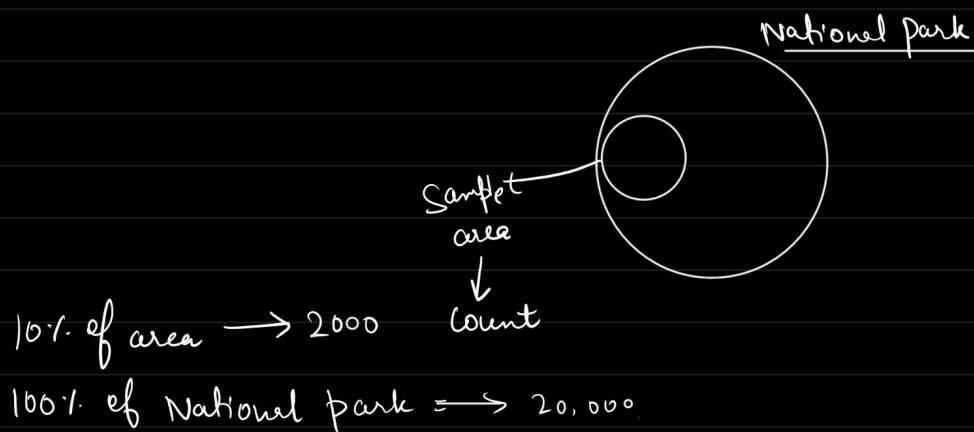
② Inferential Statistics

Why? \rightarrow Population is large
 \rightarrow time and resource constraints

Example-1 Avg age of Population of India.



Example-2 - No of trees in a National Park



Descriptive Question

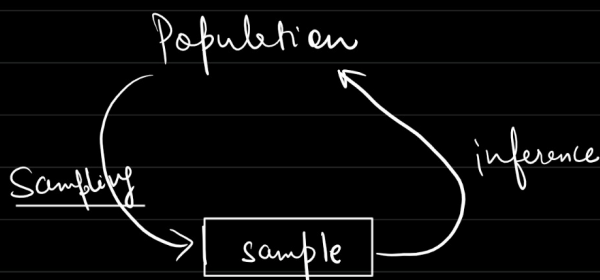
* Classroom

\rightarrow What is Avg age/height of students in class?

Inferential Question

\rightarrow 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.

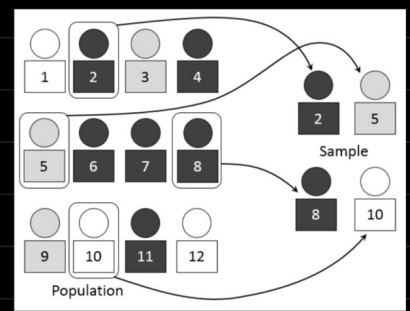
Sampling technique

- 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.



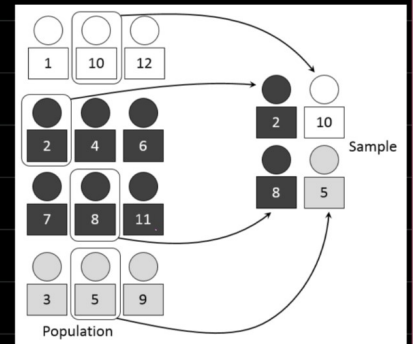
② Stratified Sampling

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.

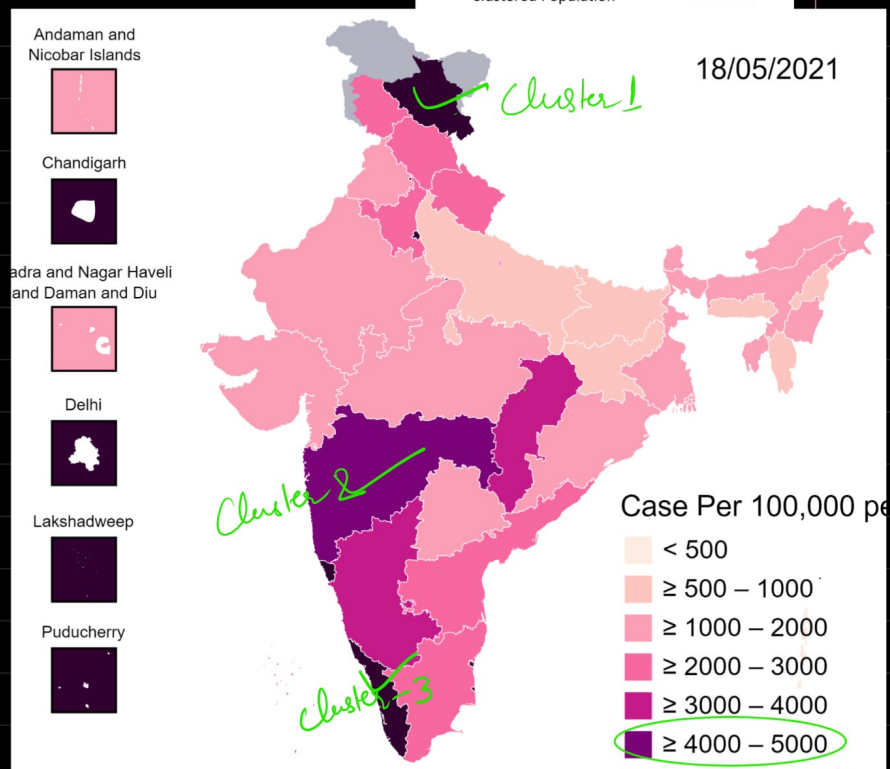
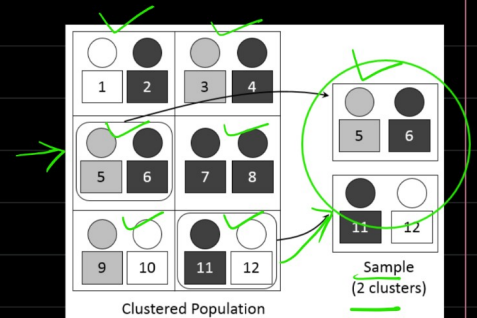
→ A simple sample would be chosen from each strata or layer.



③ Cluster Sampling

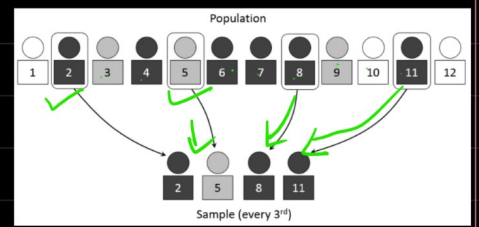
→ divides the population into groups or clusters, some of these clusters are randomly selected.

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



④ Systematic Sampling

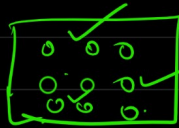
→ Every n^{th} element will be selected.



Examples → odd roll no
→ People born on odd years.

Recap

→ SRS



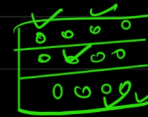
→ Quota Sampling
→ Min-max Sampling
→ Convenience Sampling
→ Accidental Sampling

→ Stratified Sampling

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

eg. Population of India

eg. Gender — Male / Female



→ Cluster Sampling

All the data



of randomly selected clusters as sample.

→ Systematic Sampling

every n^{th} member

ex. Students odd roll no

eg. people born in leap years

