Introduction to Statistics

* **Defⁿ**: Statistics is the science of collecting, organizing and analyzing data.
* **Data**: "facts or pieces of information."
* **eg**: 1) Height of students in classroom
  2) No. of sales in terms of revenue of a company.
  3) Salary of people in a society.
  4) IQ of students in classroom.

* Types of Statistics

1) **Descriptive Statistics**

Defⁿ: It consists of organizing, summarizing and visualizing data.

1) Measure of Central Tendency
   (Mean, Median, Mode)
2) Measure of Dispersion
   (Variance, S.D., Z-score)
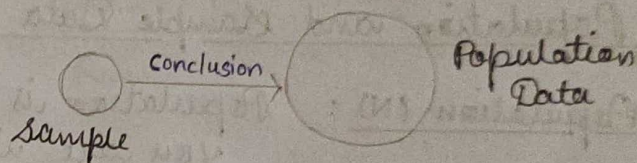3) Different types of distribution of data.
   eg: Histogram, P.d.f., pmf, Gaussian distribution, Log. Normal dist., Exponential, Binomial, Bernoulli, Dist, Poisson.

} EDA & FE

Plots

Box plot, Violin plot, Histplot, Distribution plot.

EDA → Exploratory Data Analysis
FE → Feature Engineering

2) **Inferential Statistics**

Defⁿ: It consists of using data you have measured to form conclusion.

Sample — Conclusion → Population Data

1) Z-test
2) t-test
3) chi-sq. test
4) Anova test

Hypothesis testing, P-value, significance value.

Eg: Let's say there are 10 cricket campus in Banglore and you have collected the height of cricketers from one of the camp.

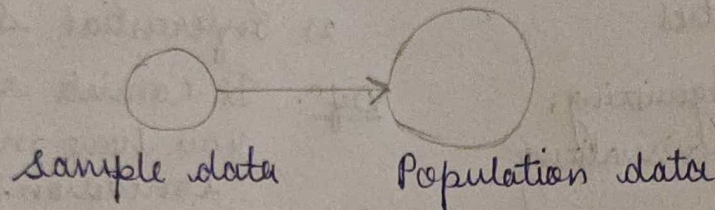**Ans :** Heights are recorded [175 cm, 180 cm, 140 cm, 140 cm, 135 cm, 160 cm, 135 cm]

SAMPLE

Descriptive Question :
- what is the average height of the entire classroom?
- Distribution of a Data.
- 140 cm how many S.D. it is away from mean. $\longrightarrow$ Z-score

  Data point

Inferential Question :
Are the average height of players of camp1 similar to that of camp 2.

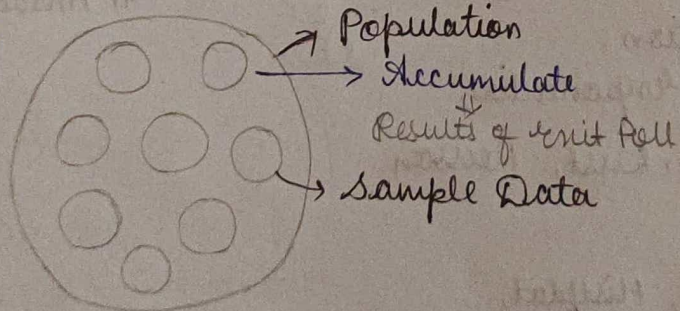Sample data $\longrightarrow$ Population data

* Population and Sample Data

Population (N) : Population is a group or a superset of data that you are interested in studying.

Sample (n) : A sample is a subset of population data.

example: Exit Poll $\rightarrow$ News channel

$\downarrow$
Outcomes

$\rightarrow$ Population
$\rightarrow$ Accumulate
Results of exit Poll.
$\rightarrow$ Sample Data

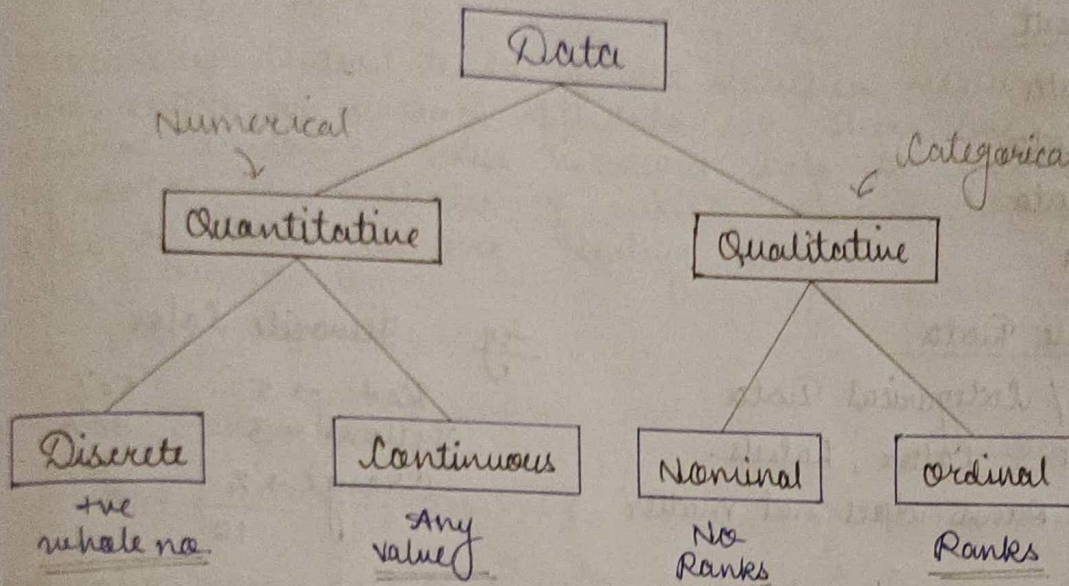| Party A | $\rightarrow$ 56 |
| Party B | $\rightarrow$ 30 |

Survey of Data Science $\underset{send\ to}{\Longrightarrow}$ Tech people

Example: Types of sampling Techniques (Assignment)

* **Types of Data :**

House Price Data

Size	Area	No. of Room	Location	Price
1500	Posh	3	New York	—
—	—	—	Banglore	—
			New Delhi	

```
                        ┌──────┐
                        │ Data │
                        └──────┘
         Numerical ↓            ↘ Categorical
      ┌──────────────┐         ┌─────────────┐
      │ Quantitative │         │ Qualitative │
      └──────────────┘         └─────────────┘
       /          \             /          \
┌──────────┐ ┌────────────┐ ┌─────────┐ ┌─────────┐
│ Discrete │ │ Continuous │ │ Nominal │ │ Ordinal │
└──────────┘ └────────────┘ └─────────┘ └─────────┘
   +ve          Any            No          Ranks
  whole no.    value          Ranks
```

Examples :

Discrete	Continuous	Nominal	Ordinal
No. of children in a family.	Average rainfall	Colour	Customer Feedback (ratings)
No. of Accounts in banks.	Length of River	Cities	Weekdays
No. of bikes.	House price in Banglore	Country	
No. of people working in a Company.		Blood group	
		Location	
		Days	
		Pin Code	

* **Example Assignment :**

1) Length of diff. rivers in the World ?
2) Favourite food based on Gender ?
3) Marital Status ?
4) IQ Measurement ?
5) Weekdays for IT employees ?
6) Weekdays for Krish ?