Statistics

Statistics: It plays a crucial role in data science by providing methods to analyze and interpret data.

collecting, organizing and analyzing

Decision making process

Data → Pieces of Information / Facts

eg → Employee salary in a company

= { 40,000, 50000, 60000, 70000, 20000)

→ Heights of students

{ 160cm, 180 cm, 190cm, 175cm ---- }

Types of statistics

(1) Descriptive (2) Inferential

↓

summarizing data

(1) Measure of central Tendency
(Mean, Median, Mode)

(2) Measure of Dispersion [variance, std]

(3) Distribution
pdf, pmf, histogram

II) Inferential

It consists of using of data you have measured to form conclusion or make a predictions about population based on a sample data.

→ Hypothesis Testing, Confidence Interval
Regression Analysis, ANOVA, Chi-square

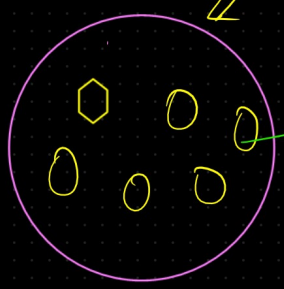
Population Vs Sample

Population: Collection of all items of interest. [1000]

(N)

Sample: A subset of the population (100)

(n)



Exit Poll

(1) less time consuming
(2) less costly

Types of data

- Categorical
  - Car Brands
    - TATA
    - Audi
    - BMW
- Numerical
  - Discrete
    - finite number (whole number)
  - continuous
    - infinite and Impossible to count

↓
No. of children

↓
weight, height temperature

Measurement Levels

Qualitative

Nominal
↓
categories
- sex
  - ↓
  - Male / Female
- Blood Group

Ordinal
↓ RANK / strict order

Customer Feedback
- Good
- Bad
- Better

Quantitative

Discrete
⇓
whole number

Continuous
⇓
decimal
Any values

Scale of Measurement

(1) Nominal Scale Data
↳ categories

(2) Ordinal Scale Data
↳ Rank

A
B
C

(3) Interval Scale Data

0 – 10] = 10
10 – 20] = 10
20 – 30] = 10
30 – 40]

(20)

(4) Ratio Scale Data

20, 40, 60, 80

$$\text{Ratio} = \frac{40}{20} = (2:1)$$