Statistics

Statistics is a field that deals with collection, organization, analysis, interpretation & presentation of data

↓

decision making.

→ Understand the data & make some decision.

↓

grow your business.

eg: Bank → 5km → B
loc. A
ATM
ATM.

How much crowd takes place and at which place, acc to that bank will see where should ATM established.

Applications
1. Everyone uses it
2. ML
3. Data analyst
4. Business analyst
5. Risk analyst.

} In every domain, statistics used.

Types of statistics
- Descriptive
- Inferential.

Descriptive
It consist of organizing & summarizing of data.

Inferential.
collect data → sample data
↓
make conclusions & influences using some experiments (z test, t test)
↓
other data
population data.

pop. data size > sample data size.

1. Measure of central Tendency
→ mean, median, mode
2. Measure of dispersion
→ variance
→ standard deviation

↓ population data.

eg: college A → 1000 students

Class stats

height of stu = { 180cm, 162 cm, 170 cm, 150cm } → sample data.

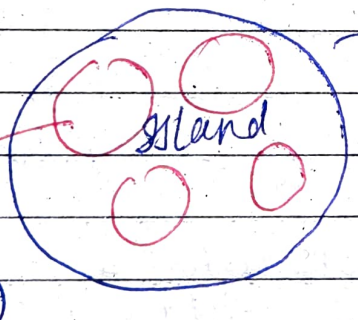
Descriptive → Mean, Median, Mode.

Inference → by seeing height of 4 student, predict the height of 1000 students.

↓

conclusion / Inferences.

Population and sample



→ 100k people

↓

population data (N)

10k

↓

sample data (n)

Task :- collect all the weights of all the people.

Solu → It is difficult to go and ask weight ~~of~~ ~~one~~ one by one to all people.

Then we will take weight of some people (10k) known as sample data.

→ showing the central part of distribution.

Date
Page

Measure of central Tendency

1. Mean
2. Median
3. Mode.

$$\mu = \sum_{i=1}^{N} \frac{x_i}{N}$$

mean of population data.

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$$

mean of sample data

eg Age = {1, 3, 4, 5}

$$\mu = \frac{1+3+4+5}{4} = \frac{13}{4} = 3.25$$

② Median :- Ages = {1, 3, 4, 5, 100} → outlier

$$\text{mean} = \frac{1+3+4+5+100}{5} = 22.6.$$

★ by adding 100, means to 22.6 from 3.25

To decrease the impact of that outlier, we use median.

Median

first sort {1, 3, (4, 5) 100}

$$\frac{4+5}{2} = 4.5$$

avg = 3.25 → just median (4.5)

by adding just 100.

→ just a small change we get.