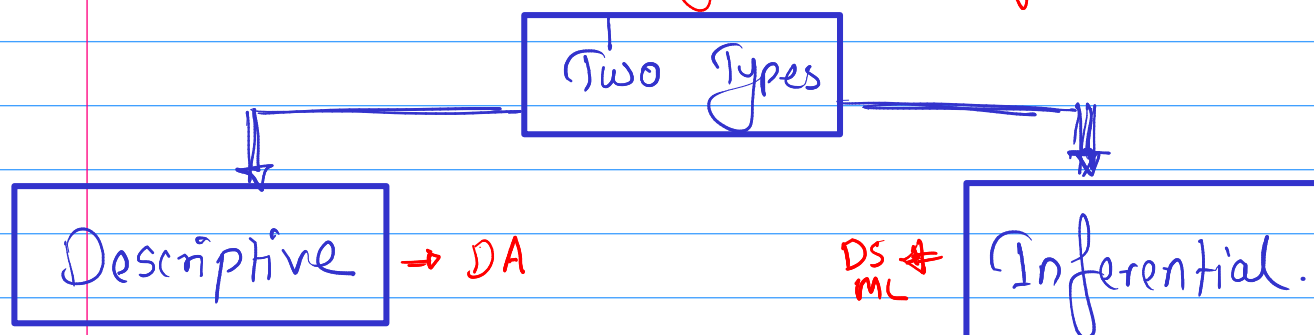


Statistics \Rightarrow It is a branch of Math,

which deals with data Collection, organise, analysing, Interpretation & presenting the data.

Data \Rightarrow Its a fact or Information



It deals with data Collection, organize, interpretation, Analysing & presenting the data.

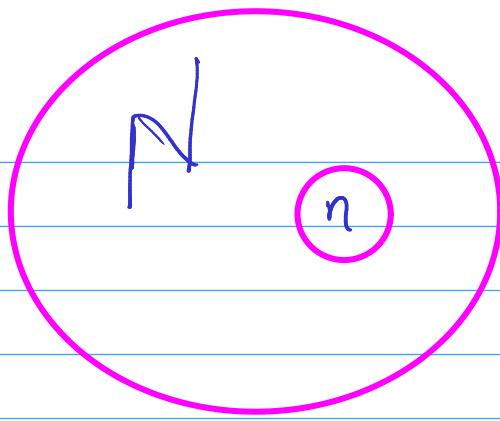
It deals with making a conclusion or the prediction about the population data with the help of sample data.

Sample $\Rightarrow (n)$ \Rightarrow It is the subgroup of the population.
 \uparrow
small N

Population $\Rightarrow (N)$ \Rightarrow It is the entire group.

Ex \Rightarrow The fan following of the cricket in the world. (Population)

The fan-following at the stadium (Sample)



Type of Data.

Qualitative / Categorical

Nominal

No-order / NO Rank
ex - Male / Female

ex - Caste
Category

Ordinal

ex - Feedback of
Resto.

Bad - 1★
good - 2★
better - 3★
Best - 4★
Excellent - 5★

Crack % → Gn, SL, SAC,
QAC, IAC

price % → 1st, 2nd, 3rd.

Quantitative / Numerical

Discrete ✓

Whole no

① no. of childrens
1, 2, 3, 4, 5, ✓

② Marks of student
0, 2, 2, 3

③ Roll No.

Continuous

① Height.

② weight.

③ BP

④ KM

Statistics provide us tools & measures to understand the data.

① Measure of Central Tendency

It is a Statistical measure that represents a typical or Central value for a dataset.

① Mean \Rightarrow It is the average Value.
Sum of all Number divided by total no.

② Median \Rightarrow It the middle value of dataset after Sorting.

③ mode \Rightarrow It is the repetitive Value in dataset.

\longrightarrow

Mean \Rightarrow $[1, 2, 3, 4, 5]$

$$= \frac{1+2+3+4+5}{5} = \frac{15}{5} = \boxed{3}$$

Note \Rightarrow It is Robust to outlier

The unwanted value in the dataset.

$[1, 2, 3, 4, \boxed{5}, \boxed{15}] =$

$$\frac{1+2+3+4+5+15}{6} = \frac{30}{6} = \boxed{5}$$

§ \Rightarrow If the outlier is in the data we do not use the mean as a center value.

② median \Rightarrow

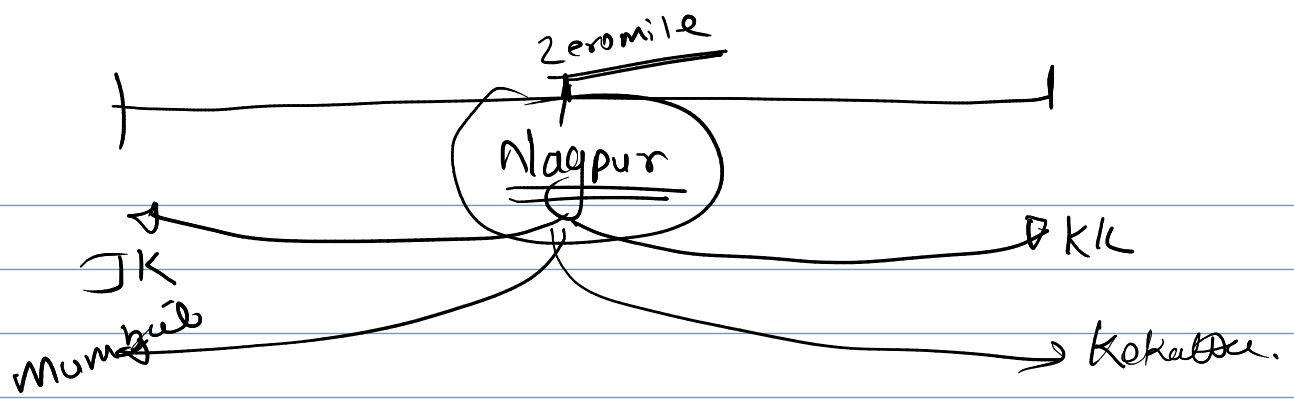
[1, 2, 3, 4, 5, 66666]
66666, 5, 4, 3, 2, 1
3.5

Mean / median \Rightarrow We use this to replace the missing value in the dataset.
We use this for the numerical data.

ex = Graduation class Age
(20, 21, 22, 23] $21.5 \rightarrow$ [21]
100 Student data \rightarrow Age] 10 Age missing

mode \Rightarrow We use mode to replace Categorical data.

100 student data - Gender \rightarrow [Male = 65] ✓ mode
Female = 20
[missing - -]
M = 45] [M = 46]
F = 45] F = 45



② Measure of Dispersion

Spread

It describe us the spread or Variability of a dataset from the mean.

① Range % \Rightarrow Max - Min

It Robust to outlier.

② Varianc % \Rightarrow It is the average of the Squared difference between each data point and the mean.

$$\sigma^2 = \text{Varianc} = \sum_{i=1}^N \frac{(x_i - \mu)^2}{N}$$

Sigma²

$$\sigma^2 = \sum_{i=1}^n \frac{(x_i - \mu)^2}{n}$$

data = [1, 2, 3, 4, 5]
 $x_1 \ x_2 \ x_3 \ x_4 \ x_5$

$\mu = \text{mean} = \frac{1+2+3+4+5}{5} = \frac{15}{5} = 3$

ex $(x_i - \mu)$ & $(x_i - \mu)^2$

$\mu = 3$

1 $(1-3) = -2 \rightarrow 4$

2 $-1 \rightarrow 1$

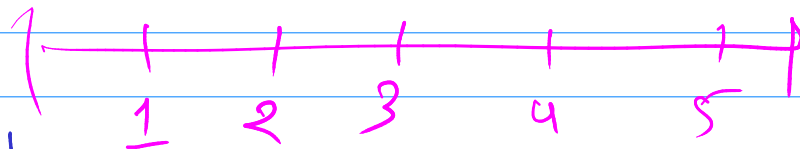
3 $0 \rightarrow 0$

4 $1 \rightarrow 1$

5 $2 \rightarrow 4$

$\frac{4+1+0+1+4}{5} = \frac{10}{5} = 2$

$\sqrt{2} \rightarrow 3 \rightarrow 2$

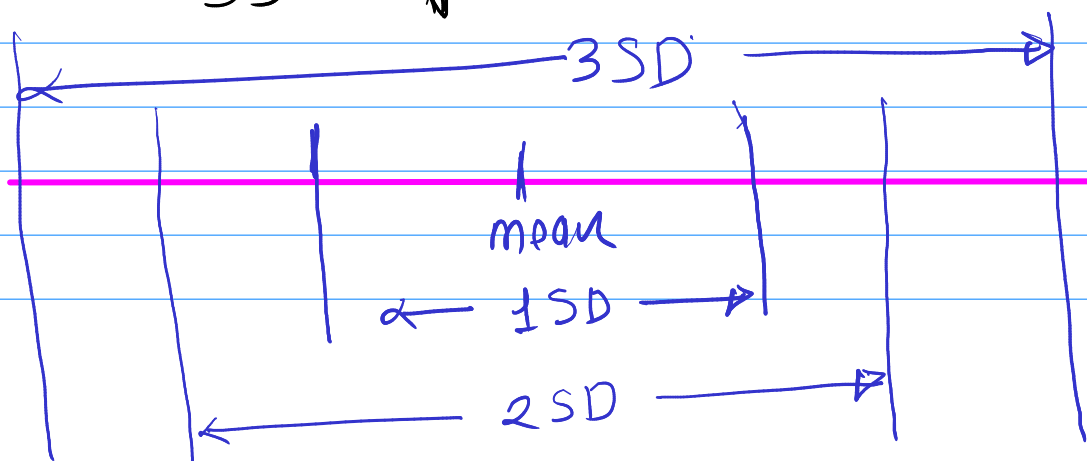


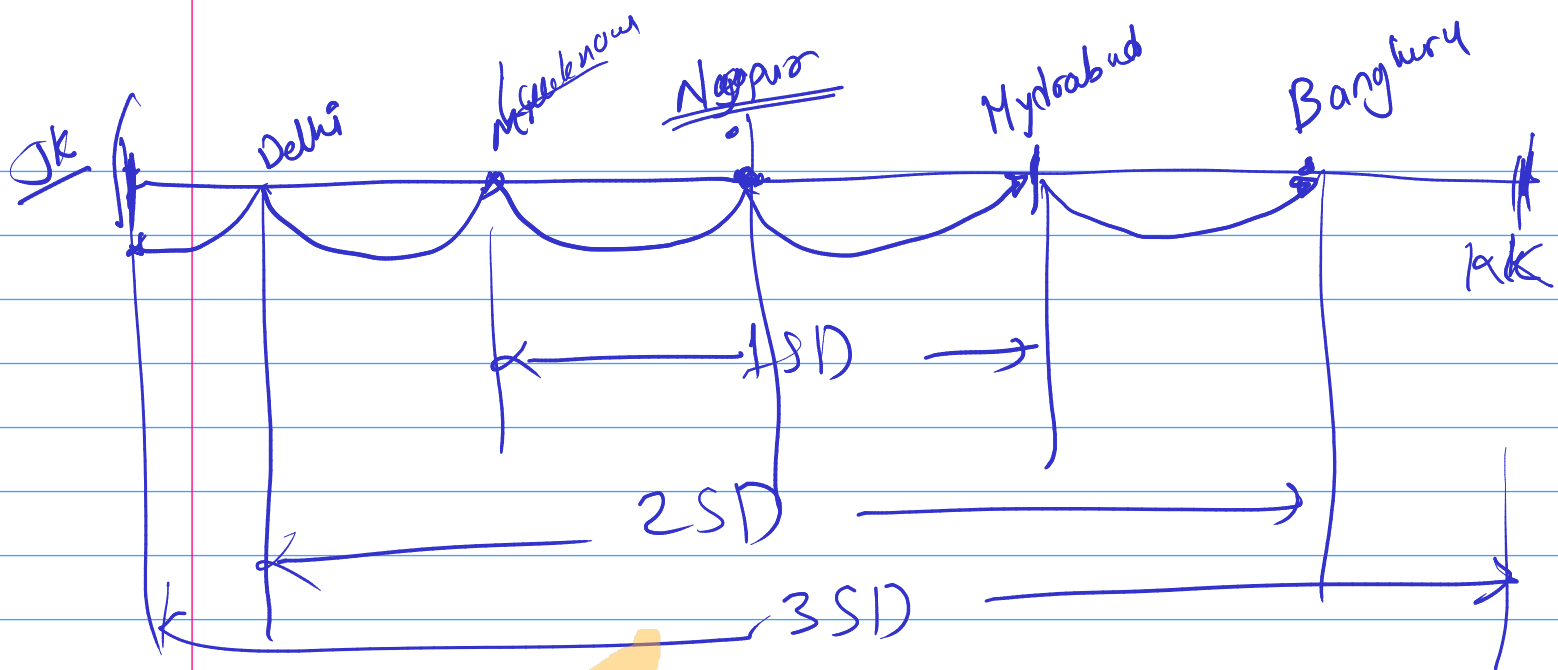
km^2

③ Standard Deviation (SD)

\Rightarrow It is the square root of the variance.

$$SD = \sqrt{\sigma^2}$$





④ Coefficient of Variation (CV)

- It is the ratio of Standard deviation & the mean.

$$CV = \left(\frac{SD}{\text{mean}} \right) \times 100\%$$

Ex \Rightarrow Biology, Chemistry, Engg.

Year of exper

1 year
2
3
4

Salary in Lakhs

1 Rupees
2
3
4

Graphs for Analysis

Clonix
1 mg

Cipla
1 g