

Statistics \Rightarrow It is a branch of Math,

which deals with Data Collection, Organise, analysing, Interpreting & Presenting the data.

Data \Rightarrow Its a fact or Information

Two Types

Descriptive \rightarrow DA

DS
ML

Inferential.

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

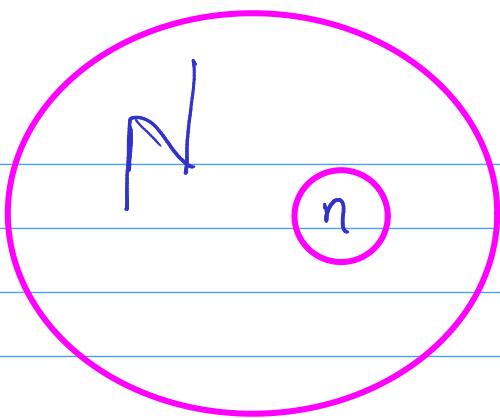
It deals with making a conclusion or the prediction about the population data with the help of sample data.

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

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

Ques \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- Castle
Category

ordinal

no-Feedback of
Resto.

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

Train ⇒ GN, SL, 3AC,
2AC, 1AC

Price ⇒ 1st, 2nd, 3rd.

Quantitative /
Numerical

Discrete ✓
Whole no

① no. of childrens
1, 2, 3, 4, 5, ✓

② marks of student
0, 2, 3

③ Roll No.

④ Height.

⑤ weight.

⑥ BP

⑦ 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 :- It is the average value.
Sum of all number divided by total no.

② Median :- It is the middle value of dataset after Sorting.

③ mode :- It is the repetitive value in dataset.

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

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

Note :- It is Robust to outlier

The unwanted value in the dataset.

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

⑧ \Rightarrow If the outlier is in the data we do not use the mean as a center value.

⑨ median \Rightarrow

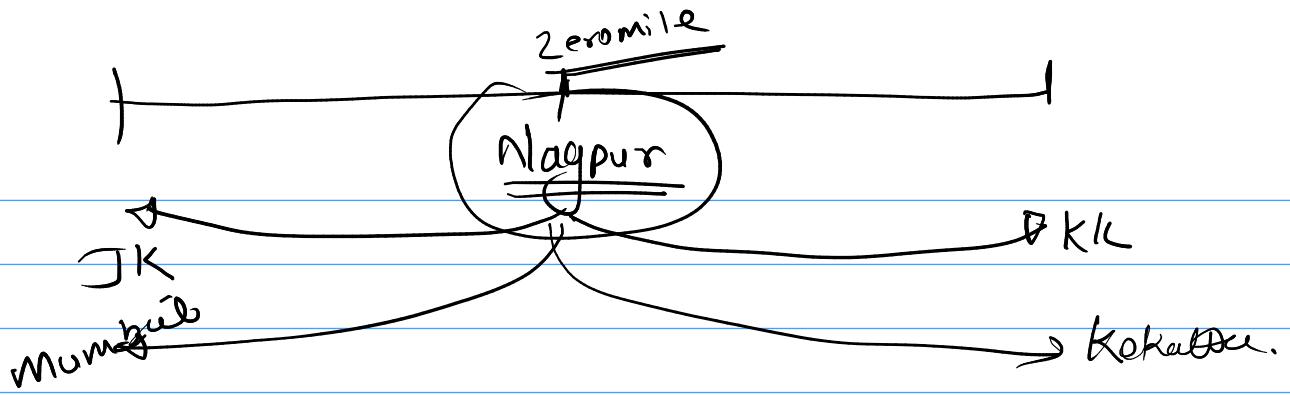
$$\begin{array}{c} [1, 2, 3, 4, 5, 6666] \\ \textcircled{6666}, 5, \underline{\sqrt{3}}, \underline{\frac{3}{2}}, 2, 1 \\ \underline{\underline{3.5}} \end{array}$$

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

e.g = Graduation class Age
 $(20, 21, 22, 23)$ $21.5 \rightarrow \boxed{22}$
100 student data \rightarrow Age] to Age missing

mode \Rightarrow We use mode to replace Categorical data.

100 student data - Gender \rightarrow Male = 65 + mode
Female = 20
 $M = 45$] $m = 46$] $F = 45$]
missing - -



② Measure of Dispersion Spread

It describes the spread or variability of a dataset from the mean.

① Range : $\Rightarrow \underline{\text{Max} - \text{Min}}$

If Robust to outlier.

② Variance : \Rightarrow It is the average of the squared difference between each data point and the mean.

$$\sigma^2 = \text{Variance} = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}$$

~~σ^2~~ σ^2

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

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

$\mu = \text{mean} =$

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

σ^2 $(x_i - \mu)^2$ $(x_i - \mu)^2$

$\boxed{\mu = 3}$

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

2 $-1 \rightarrow 1$

3 $0 \rightarrow 0$

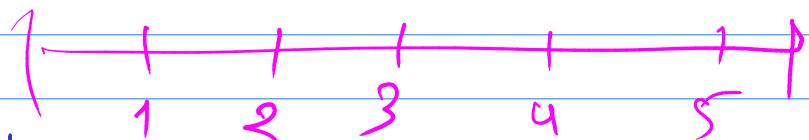
4 $1 \rightarrow 1$

5 $2 \rightarrow 4$

$4+1+0+1+4$

$= \frac{10}{5} \rightarrow \boxed{2}$

$\star \boxed{3} \star$

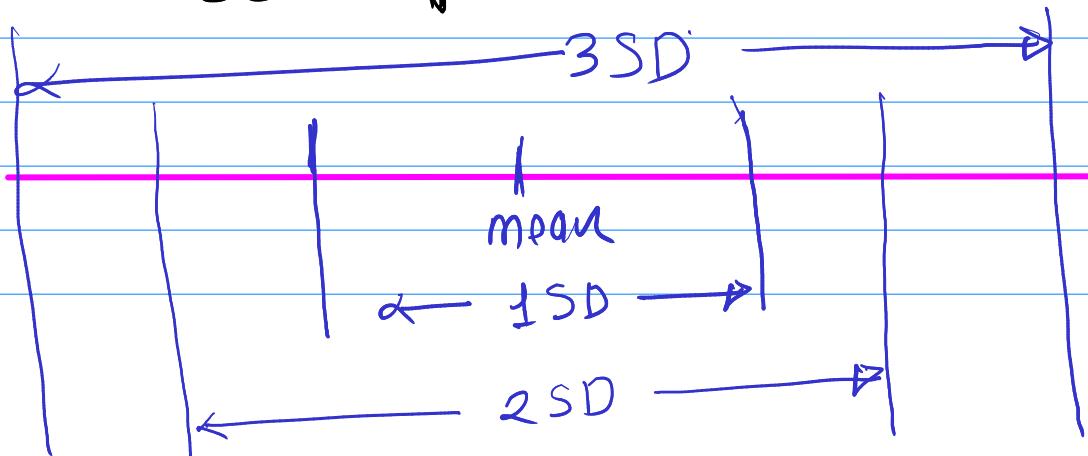


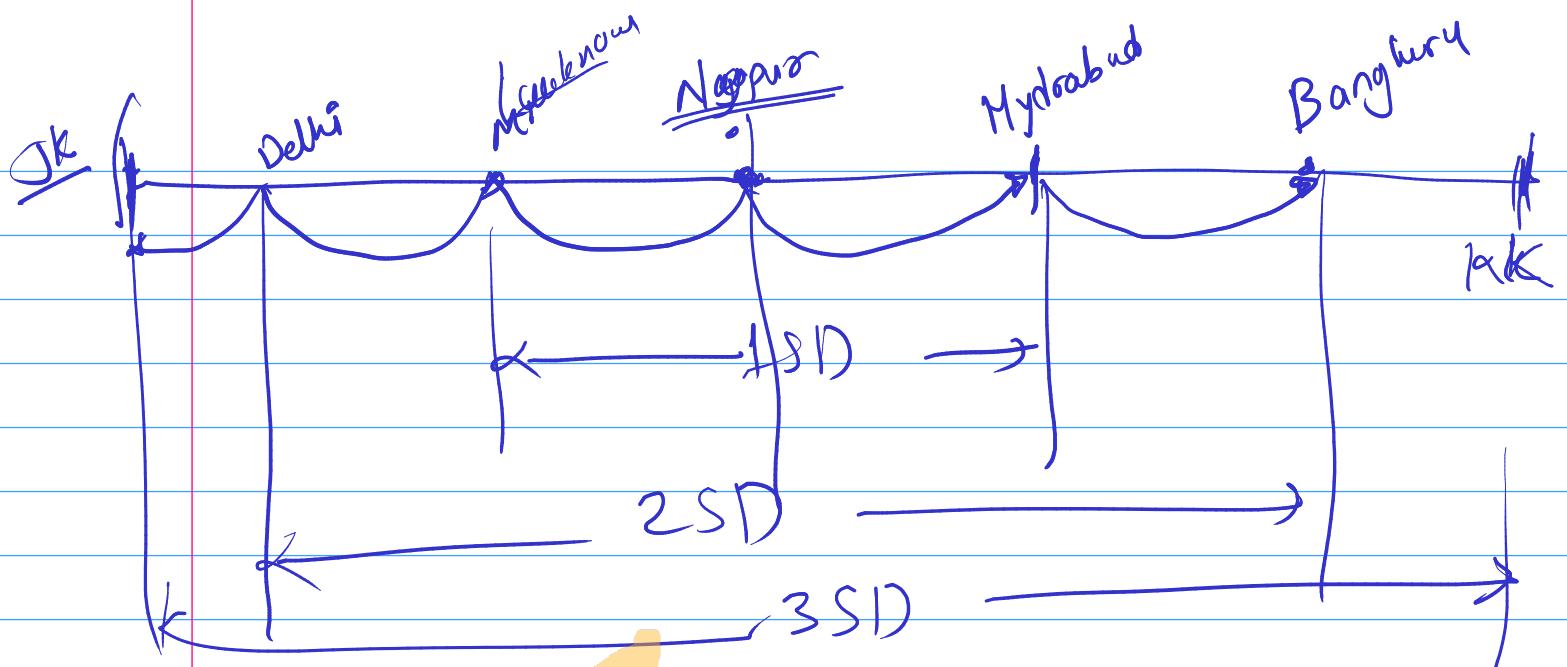
$(km)^2$

③ Standard Deviation (SD)

⇒ 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.

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

e.g. \Rightarrow Biology, Chemistry, Engg.

Year of exper

1 year
2
3
4

Salary in Lakh

1 Rupees
2
3
4

Graphs for Analysis

Clonix | Cipla
1 mg | 1 g