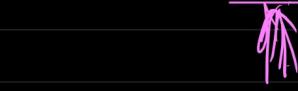


Statistics

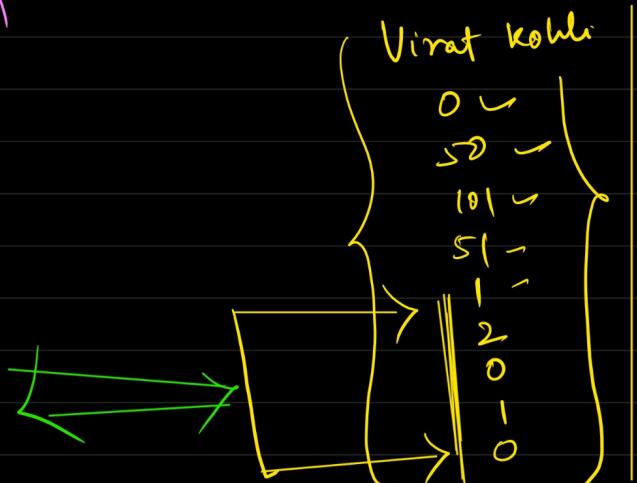
→ Statistics is a part of math which deals with collecting and analyzing of data.

→ Statistics is a mathematical science including methods of collecting, organizing and analysing data in such a way that meaningful conclusions can be drawn from them.

data → data is a piece of information that can be stored and retrieved.



Rahul Dravid



Motivation of Statistics

→ Weather forecast (meteology)

→ Sports Analytics.

→ Election Campaign.

→ FMC () E-commerce

→ medical / Genetics

→ Stock market

2005

↳ Cyclone → 10k+

(IPAC)



Each of the Constituency

1999 — 2020



CovidIndia.org

Map of India

types of Statistics

↓
Descriptive

↓
Describe

→ organizing and summarizing
the data | population

Inferential
Statistics.

↓
Inference.

Virat Kohli :- [100, 0, 10, 5, 100, 10]

Arg Score / Strike rate.

→ 150 cm 160 cm 180 cm 150 cm 180 cm

What is avg ht of students in class?

① Measure of Central tendency
(mean, median, mode)

② Measures of Spread
(Std dev, Variance)
(Range, Percent, Decile, Quartile, IQR, Box Plot)

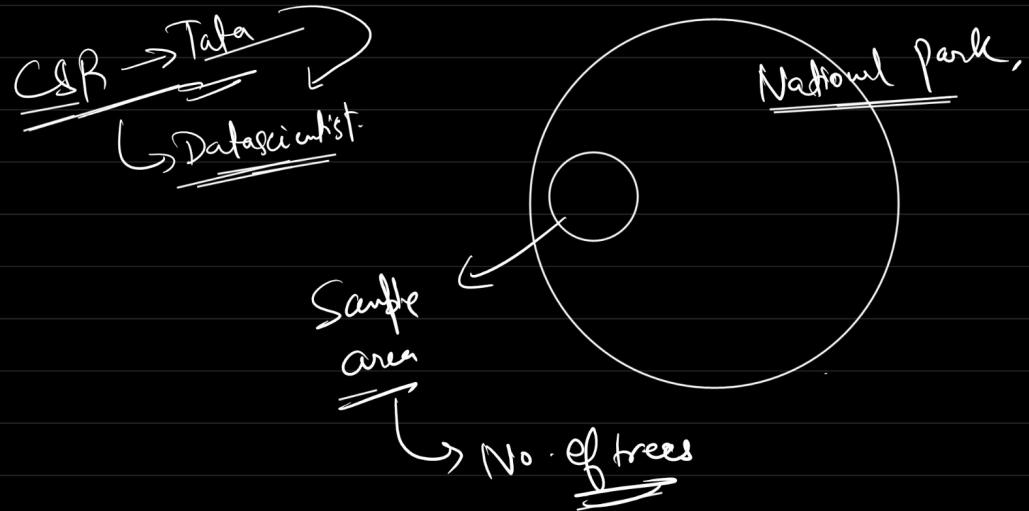
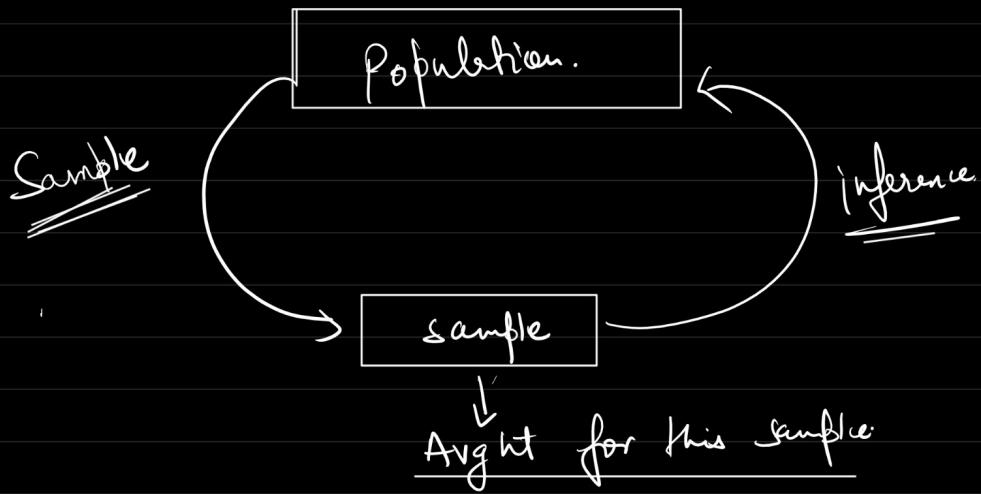
③ Measures of Symmetry
(Skewness, Kurtosis)

* Inferential Statistics

What is the avg ht of population of India ??

→ Resource
→ Time.

→ Population is huge



Population → Whole data

Sample → Subset of Population Area

Avg ht of population of India

Q. J. H. M. H. T. N. P. J. H. V. P.

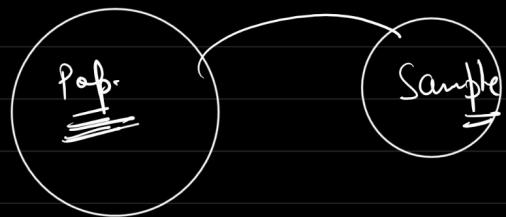


⇒ Avg ht → 6.1 ft

Sikkim
Assam
Nagaland
Meghal.
Tripura
Mizoram

Sampling Techniques

- [1] Simple random sampling.
- [2] Stratified sampling
- [3] Cluster Sampling
- [4] Systematic sampling.



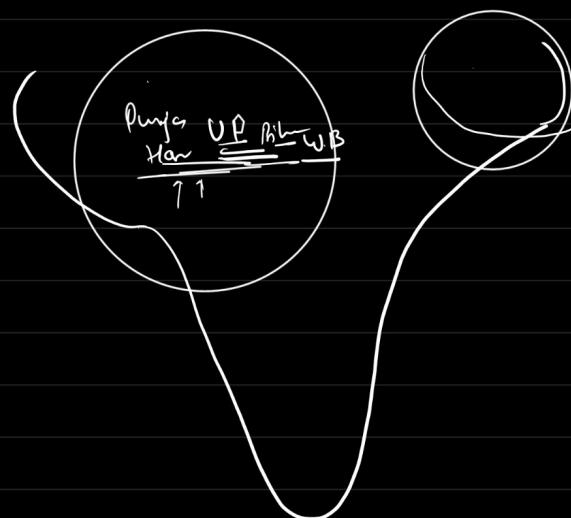
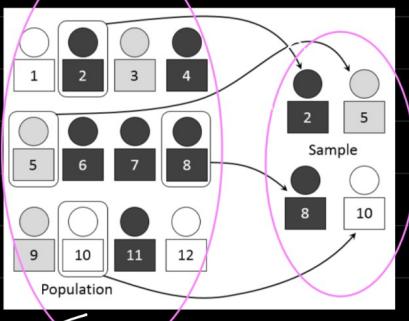
① Simple random sampling

random

→ Each member of the Population (N)

has an equal chance of being selected for your sample.

$$\left[\begin{array}{c} \textcircled{0} \\ \textcircled{0} \\ \textcircled{0} \\ \textcircled{0} \\ \textcircled{0} \end{array} \right] \Rightarrow \frac{1}{5}$$

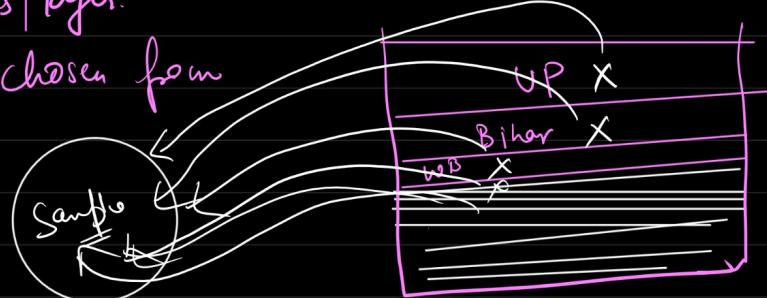


② Stratified Sampling

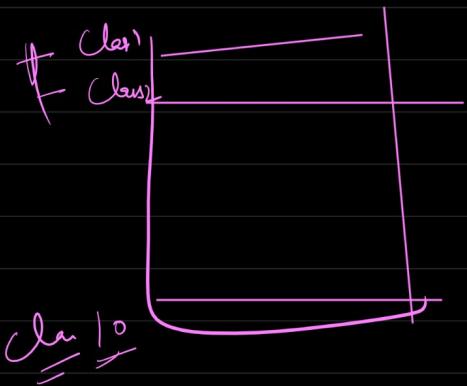
Strata \Rightarrow groups | layers

→ different distinct groups | layers.

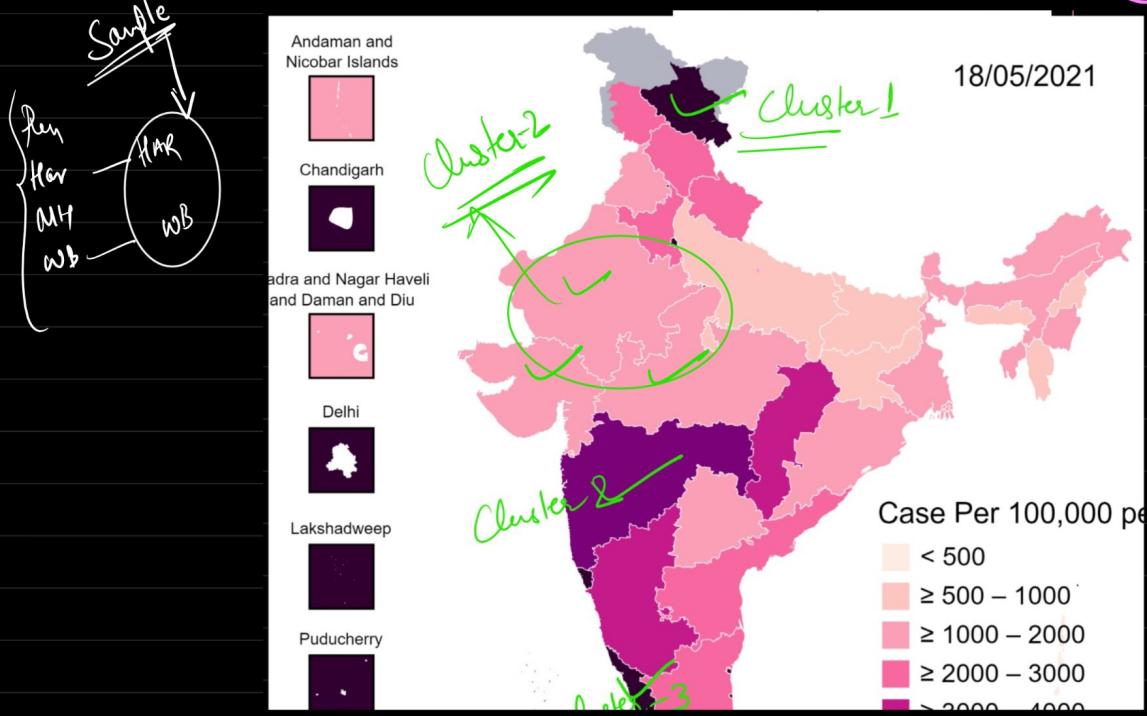
→ The elements would be chosen from each strata



Aug 1st
of
population of
School

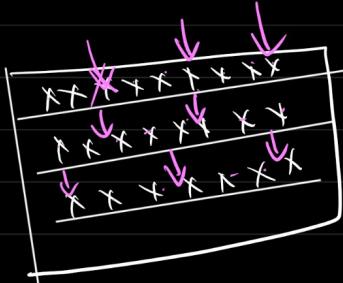


- ③ Clustering → divides the population into groups or clusters.
Some of these clusters are randomly selected.
All the individuals in the chosen cluster
are selected to be in Sample.



④ Systematic Sampling

↳ Element is selected in a systematic way.



Every n^{th} element

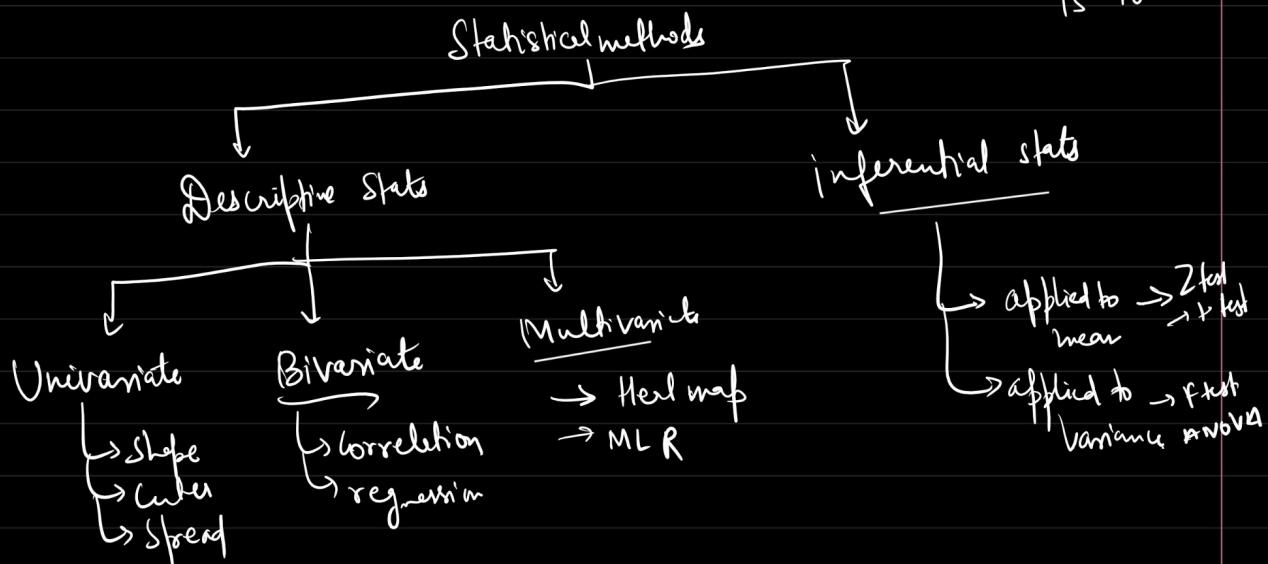
\Rightarrow odd roll no
 People born
 in leap year

Descriptive

→ Avg ht of population

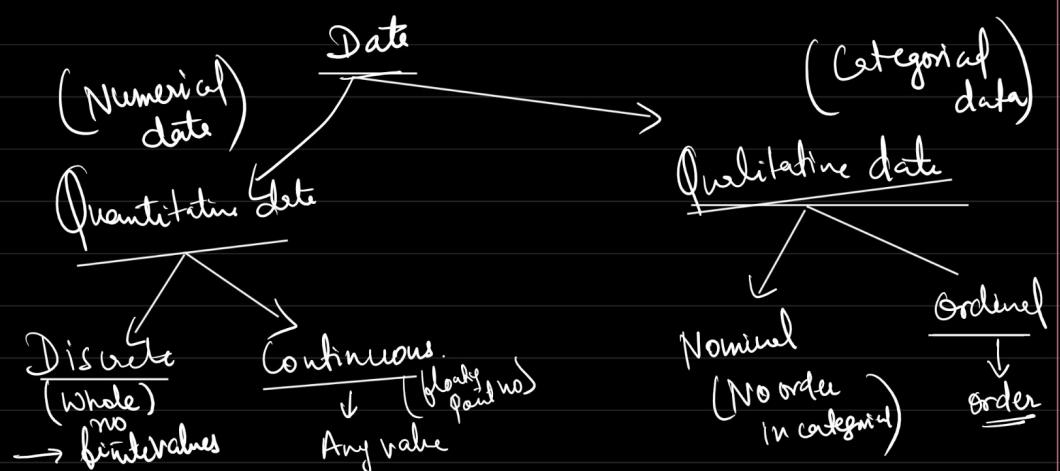
Inferential

→ If sample avg ht is 165 cm then it is true that the avg ht of population is 165 cm.



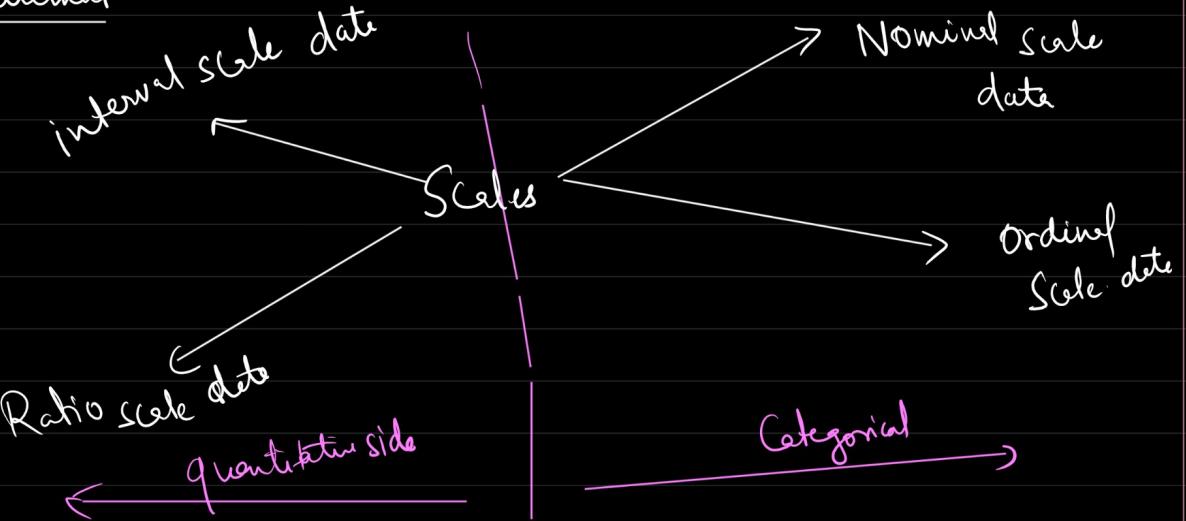
Type of data

$+/-\div$



| | | | |
|-----------------------|------------------|-----------------------|-------------------------------------|
| ① No of children | in a range | Gender | eg. Good better best |
| ② No of phones | → Height | → blood grp | |
| ③ Bank | → weight | → color | eg. Grade |
| ④ employees in a team | → Speed | → race | eg. Salary (high mean low) |
| | → temp | → marital status | |
| | → IQ | → ethnicity | |
| | → movie duration | | |
| | → Length | (Nonnatural) order | eg. Rank in my class |
| | | | eg. Height cm |

Scale of measurement



① Nominal Scale data

- Qualitative / Categorical
- Gender, ethnicity, colour, location
- No order in the data

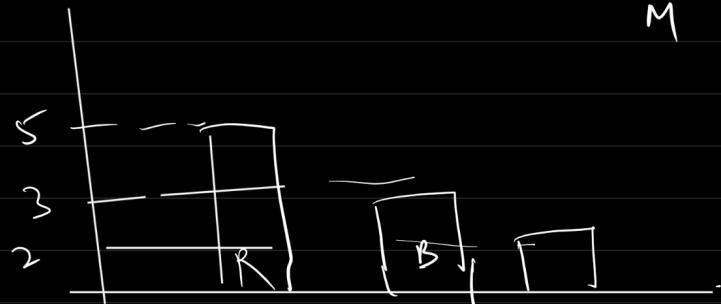
$$\begin{array}{c} \overline{\overline{M}} \rightarrow 5 \\ \overline{\overline{F}} \rightarrow 2 \end{array}$$

Employees

M
F
M
F
M
M
M

Red

Blue
Yellow



② Ordinal Scale data

- Order and rank matter
- Difference can not be measured

Performance of
Student

| | |
|----------|-------|
| Subu | — 1st |
| Rahul | — 2nd |
| Sumana | — 3rd |
| Vatsalya | — 4th |
| Agnim | — 5th |

③ Interval scale data

- The rank and order has a meaning
- rank & values both matters
- Difference can be measured
- It doesn't have a starting point

Length
width
height
temp
Review of a company

Srinivas — 58] → 3
Deep — 55] → 1
Nagendra — 54]
Mohit — 99

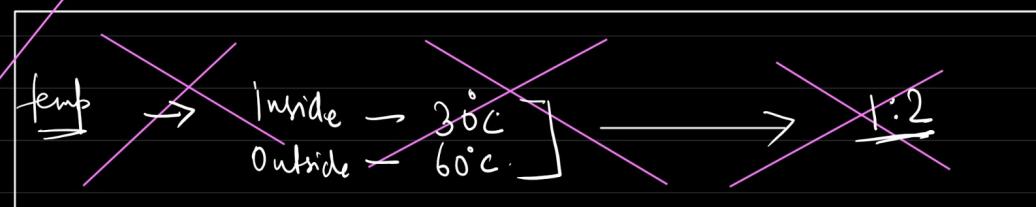
④ Ratio Scaled data

- Order and rank has a meaning
- Difference and ratio are measurable.
- It does have a 0 starting point

$$M = 50$$

$$F = 100 \quad M:F = \frac{50}{100} = \frac{1}{2} = 1:2$$

$$\underline{\text{length}} \rightarrow f = 100, A = 50 \rightarrow \frac{100}{50} = \underline{2:1}$$

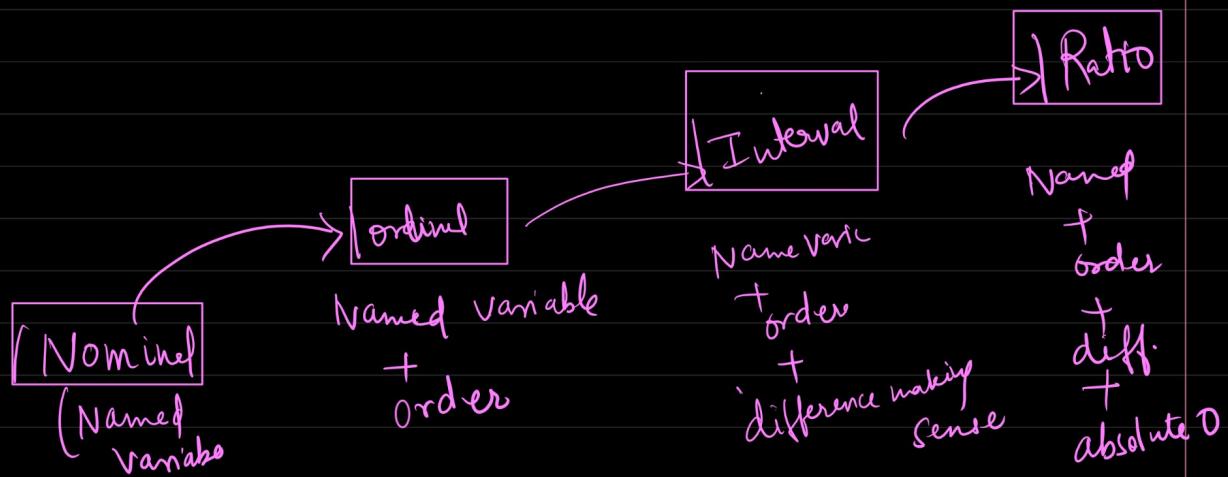


$$\text{length of Father's cloth} = 60 \text{ cm}$$

" " Son's " = 30 cm

height
but
Time
Age





Scale of Measurement

| Datu | Nominal | Ordinal | Interval | Ratio |
|-------------------------|---------|---------|----------|-------|
| Labelled | ✓ | ✓ | ✓ | ✓ |
| Meaningful | ✗ | ✓ | ✓ | ✗ |
| Measurable differens | ✗ | ✗ | ✓ | ✓ |

| True Zero Start point | ✗ | ✗ | ✗ | ✓ |
|--------------------------|--|---|---|-----------------------|
| Example | gender religion P.O code location | salification score rain force rank | Temp IQ Score ht wt length | ht wt ht age |

* Types of Analytics

- ① Descriptive Analytics → Complete Pop
- ② Predictive analytics → ML
- ③ Prescriptive analytics →

Scenario 1

Vatsalya (M)

↳ CEO → Airline Comp

→ Performance of his company

- * Avg delay of flight
- * No. of passengers
- * Revenue.

Scenario 2

VK Score

80
85
61
62

→ You predict
* Predictive Analytics

* What is revenue next year | quarter

Scenario 3

VIC Score

80

50

60

40

0

0

0

1

→ What is
prescription
by
RD

Descriptive Stats

→ Describing / summarising data without modifying it

Types

① Measure of Central tendency

mean, median, mode

② Measure of Spread / dispersion

③ Measure of symmetry.