Statistics

Statistics is the science of collecting, organizing, and analysing data.

Data ⇒ facts or pieces of information.
ex:- Ages of students in a class
$\{24, 30, 21, 34, 20\}$ ⇒ Mean, Median etc ⇒ Analysis

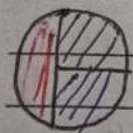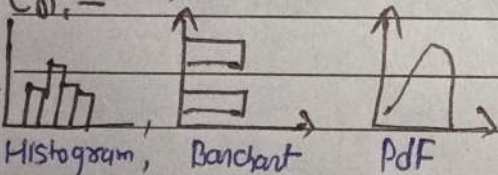ex - weights of students in a class
$\{75, 45, 50, 55, 60\}$ diagram.

Statistics (2 types)

Descriptive
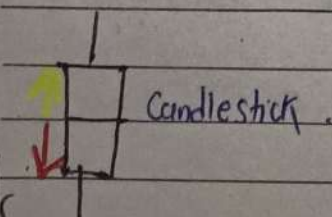
It consists of organising and summarizing of data

ex:-

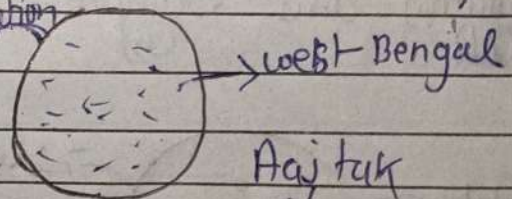
Histogram, Barchart, PDF

Piechart   Boxplot

Candlestick.

Inferential stats.

It consists of a technique by creating conclusion. to form some conclusion.
ex:- election is happening

population → West Bengal

Aaj tak
↓
Exit poll
↓
took sample of data
↓
perform some hypothesis testing
↓
Conclusion.

ex:- P-value, Z-test, t-test, chi square Annova etc.

3

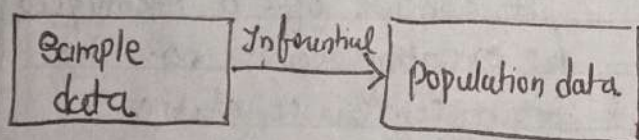★ What is the avg. size of all Sharks in the world?

#Example :- Let's say there are 20 classrooms in a university and you have collected the age of students in one classroom.

Age { 21, 20, 18, 34, 17, 22, 24, 25, 26 23, 22 }

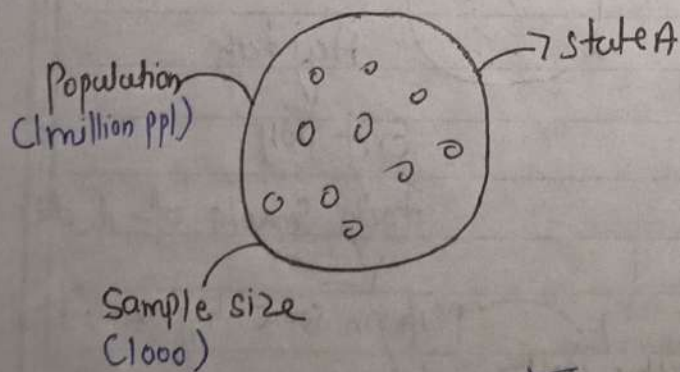Descriptive Statistics - What is the avg. age of the students in the Classroom??

Minimum age, Outliers, Distribution, Transformation, mean.age

Inferential Statistics - Are the avg age of the (Students in the ^Sample classroom) equal to the (avg. ages of the entire uni.) Population data.

| Sample data | Inferential → Population data |

Population and Sample.
(N)   (n)

eg:- Exit poll of state

→ State A

Population (1 million ppl)

Sample size (1000)

Avg. height of all the people in the State??

Sample --Avg height--> population

Sampling Techniques.

1) Simple Random Sampling - Every member of the population (N) has an equal chance of being selected for you Sample (n).

     eg-: Tossing a coin

$$Head = \frac{1}{2} , tail = \frac{1}{2}$$

     cn:- if there are 1000 people in a country then there probability of being selected is $\frac{1}{1000}$.

     ↘ ayers→ non overlapping.

2) Stratified Sampling -

     eg- Gender $\begin{bmatrix} Female \\ male \end{bmatrix}$, Blood groups $\begin{bmatrix} A+ \\ O+ \\ B+ \end{bmatrix}$

3) Systematic Sampling -

     en:- | Banglore airport |

     ↓

     | SBI credit Card stall |

(nth person selected)   O   O   (N) → Selecth every nth people.

every 5th person ← 🧍 ⟷ 🧍 → every 7th person

     A   B
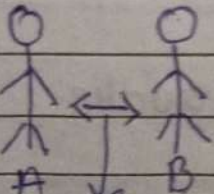
they decide how they select people

Signature...............

(5)

4) Convenience Sampling. — only those people who are intrested will only be participating.

ex - Survey regarding new technology.

Variables.

A variable is a property that can take on any value.

ex - Age = 24
Age = 25

Two kinds of variable

1) Quantitative variable.

Measured Numerically (Add, SUB, x and ÷)

2) Qualitative variable ⟹ Categorical variable

ex - Gender [female
              mual

{ Based on some charactersicas we can group categorical variable.

ex Types of flowers [rose
                     lily
                     lotus
                     sun

Quantetive Variable

Discrete	Continous.
ex - whole no.	eg - height, weight etc
0,1,2,3,4 - - .etc but	172 cm, 150,5 cm etc
not 0.1, 0.3, 1.1 etc.	

Histogram { Construct a histogram }

10.5

Ages = { 10, 12, 14, 18, 24, 26, 30, 35, 36, 37, 40, 41, 42, 43, 50, 51,
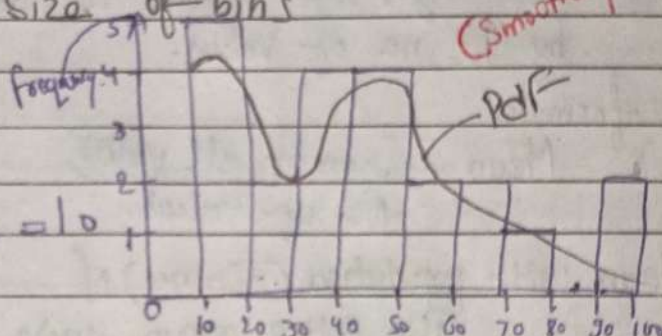65, 68, 78, 90, 95, 100 }

① Sort the no. Asending order
② Bins (nos, groups or gaps)
③ Bin size → Size of bins

ex - Suppose
[0 - 100]
Bins = 10
Bin size = $\frac{100}{10}$ = 10

(Smoothiny histo) ate
PdF

frequency

10 20 30 40 50 60 70 80 90 100

ex - [0 - 100]
Bins = 20
Bin size = $\frac{100}{20}$ = 5

5 10 15 20 25 30 35 40   100
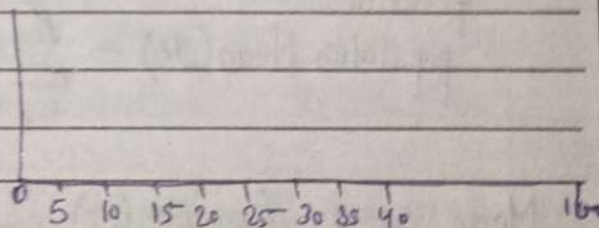
ex [0 - 100]
Bins = 5
Bin size = $\frac{100}{5}$ = 20

0 20 40 60 80 100