St. Statistics

Introduction to statistics

* To check the quality of the data.
  → Descriptive statistics.
* To make a statement or conclusion
  → Inferential statistics.

what is statistics?

It is science of collecting, organizing and analyzing data (is better decision make)

what is data?

Data means fact or piece of information that also can be measured.

Ex: the IQ of a class students

$\{98, 97, 68, 57, 110\}$

avg IQ, Min, Max

Descriptive statistics

It consists of organizing and summerizing data

Inferential statistics

Techniques where are used the data that we have measured to form conclusion. or.

To make a statement, (conclusion on a descriptive statistics

we use inferential statistics.

Q. Are the avg marks of the java class students is same as Python class students in the Besant
→ Inferential Statistics.

Q. what is the avg marks 3a class students?
→ Descriptive Statistics.

Population (N) and Sample (n)

The Entire group of data we call it as **Population**.
eg: All people in India

A Subset of population we call it as **Sample**
eg: 1 lakh people from different region of India

→ Population are larger than Sample
→ Samples should be representative of the population
→ Samples allow for easier, faster and less costly data collection.

Type of Sampling techniques :-

→ **Simple Random Sampling**

Every member of population has an equal chance being selected for our sample.
eg: The avg mileage of bike
avg ratio of marriage. Married people in Engls.

→ **Stratify Sampling**
where the population is split into non overlapping group
eg:- Person is male or female

→ **Systematic Sampling**
From the population every nth sample we will take
eg:- while doing survey in the mall on the topic of auto moderization, collecting information of every fifth person who is coming out from mall

→ **Convenience Sampling**
The sample is collected based on our convenience from the Particular domain experts.

Note Sampling technique selection always depend on problem statement.

Variable

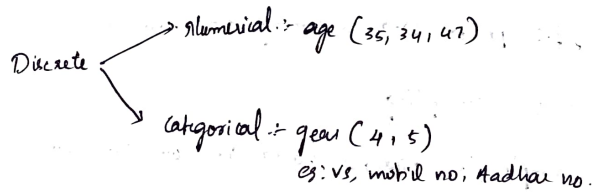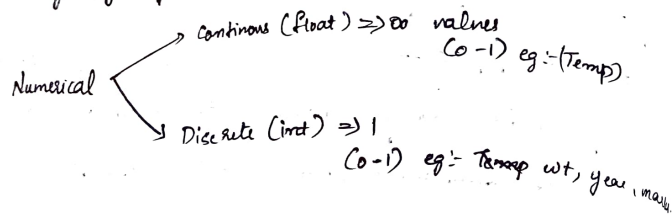A variable is a property that can take matters but natural an any value.
Two kinds of variable :-
1) Quantitative (Numerical values) variable direct, equidistant
2) Qualitative (Categorical) variable

A value can be measured and we can perform math...
operation like (A, S, M, D).

eg: mpg, weight, height

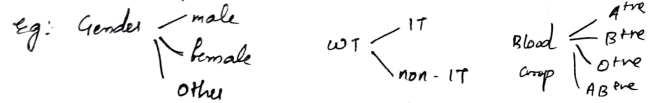Numerical
- Continous (float) ⟹ ∞ values
  (0-1) eg:-(Temp)
- Discrete (int) ⟹ 1
  (0-1) eg:- Temp, wt, year, man...

Discrete
- Numerical:- age (35, 34, 47)
- Catgorical:- gear (4, 5)
  es: vs, mobil no, Aadhar no.

→ ...

Type of Samp

→ Simple Rando
- Every memb
- being selecte
- eg: The ang
  ang rat

---

2). Qualitative / categorical variable

Non measurable data and based on some charactaristics.
we can derive catgorical variables.

eg: Gender — male
           — female
           — other

WT — IT
   — non-IT

Blood — A+ve
        — B+ve
Group — O+ve
        — AB+ve

Variable Measurement Scales

4 types of measured variables

1) Nominal Data:-
   The categorical data which one having different classes.

2) Ordinal Data:-
   Orders of the data matters but values does not

   Eg:- {50, 94, 60, 79, 40}
         4th 1st 3rd 2nd 5th

         {83, 45, 66, 96, 36}
          2nd 4th 3rd 1st 5th

3) Interval Data
   Order matters and value also matters but natural zero is not present.

4). Ratio Data
   The ratio data can be measured, ordered, equidistant and meaningful zero.