

^{30/10/24}
Note:- Sampling technique selection always depends on Problem Statement.

Different types of data and variance:-

Variable:- A variable is a property that can take an any value.

Two kinds of Variable:-

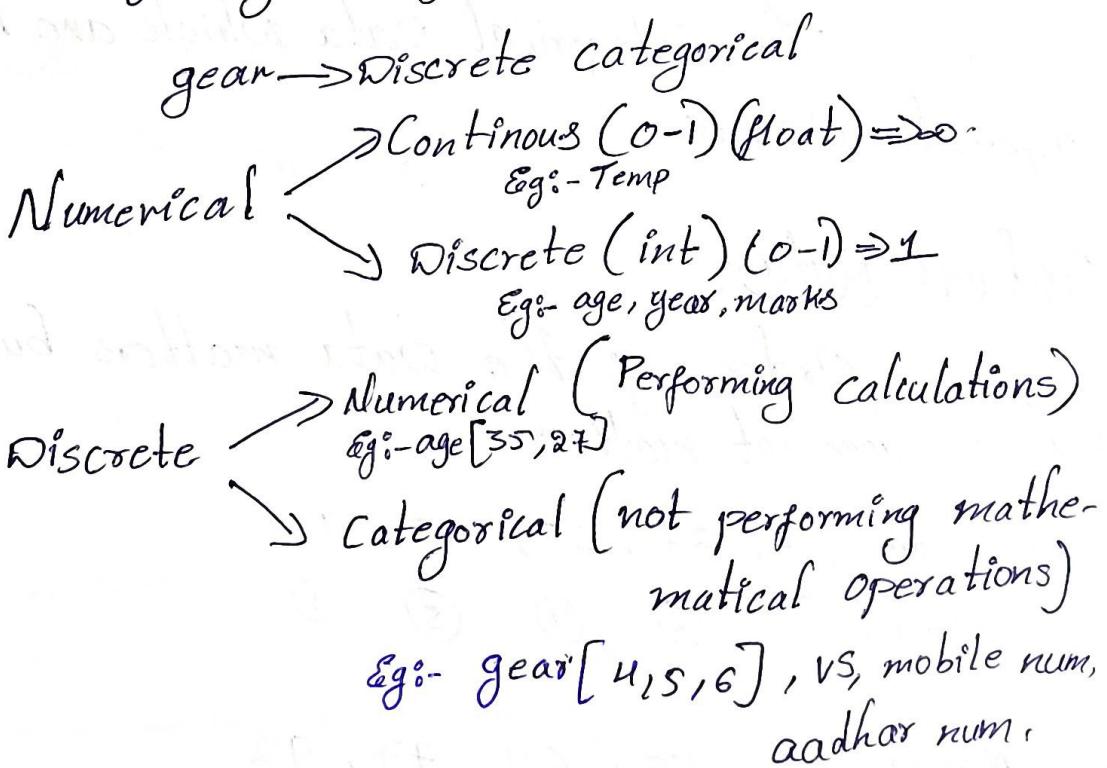
→ Quantitative (Numerical) Variables

→ Qualitative (Categorical) Variables.

→ Quantitative Variable:-

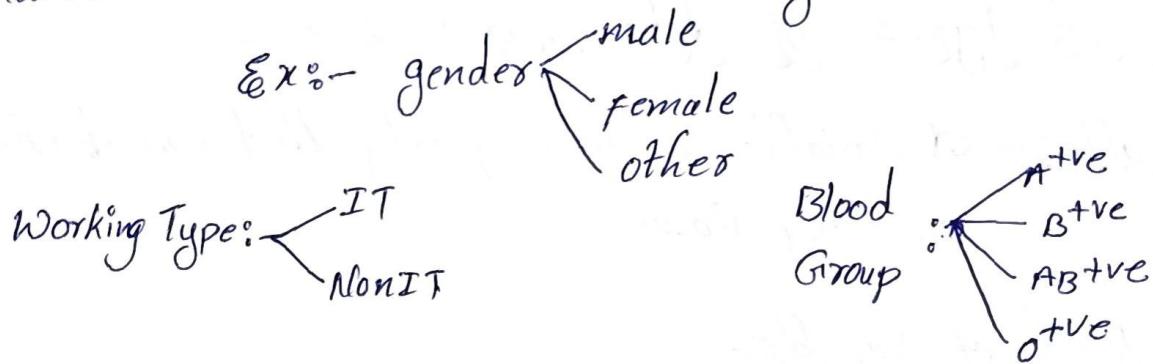
A value can be measured and we can perform mathematical operations like (Addition, multiplication, subtraction, division).

Ex:- mpg, weight, height



Qualitative Variable :-

A Non-measurable data and Based on characteristics we can derive categorical Variables.



Variable Measurement Scales :-

We have 4 types of measured variables

- Nominal Data
- Ordinal Data
- Interval Data
- Ratio

Nominal Data :-

The Categorical data which are having different classes.

Ordinal Data :-

Order of the data matters but the values does not matters

Ex:- Marks:- 67, 87, 58, 48, 97

Rank:- ③ ② ④ ⑤ ①

Marks:- 28, 45, 64, 77, 92

Rank:- ⑤ ④ ③ ② ①

Interval Data :-

Order matters and value also matters but natural zero is not present.

Ex:- Eight of Eye (it can't be zero)

Ratio Data :-

The ratio Data can be measured, Order, Equi distinct and have meaningful zeros. [two zero point]

Ex:- Ratings, no. of students, salary, height, age.

31/10/25

Descriptive Statistics :-

→ Measure of Central tendency

→ Measure of dispersion

→ Data Visualization.

Measure of Central tendency :-

→ Mean } Numerical

→ Median

→ Mode [categorical]

Mean :-

Population Mean :- (μ)

Sample Mean :- (\bar{x})

$$\mu = \frac{\sum_{i=1}^N x_i}{N} , \quad \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

↑ ↑ sum of obs
Population mean Sample Mean No of obs

Population mean

Sample Mean

Median :-

- Sort the value either asc (or) desc Order.
- choose the mid value
- If you get mid (2) values take avg of those 2 values.

Ex:- ① [1, 2, 2, 3, 4, 5], ② [1, 2, 2, 3, 4, 5, 100]

$$\text{mean} = \frac{1+2+2+3+4+5}{6} = \frac{17}{6} = 2.8 \quad \frac{2+3}{2} = \frac{5}{2}$$

$$\text{Median} = \frac{2+3}{2} = \frac{5}{2} = 2.5$$

$$\textcircled{2} \quad \text{mean} = \frac{1+2+2+3+4+5+100}{7} = \frac{117}{7} = 16.7$$

$$\text{median} = 3$$

→ Mean will be affected by outliers where as median won't affect by outliers.

Outliers :- Far Element
replacement

→ For null value Imputation we are using mean and median. [one of the use case]

Mode :- The most repeated values.

Ex:- [1, 2, 2, 3, 4, 5, 100]

2.

<u>Example :-</u>	<u>Person 1</u>	<u>Person 2</u>
Mon	7:30 AM	
Tues	7:45 AM	8 AM
Wed	8 AM	11 AM
		9 AM