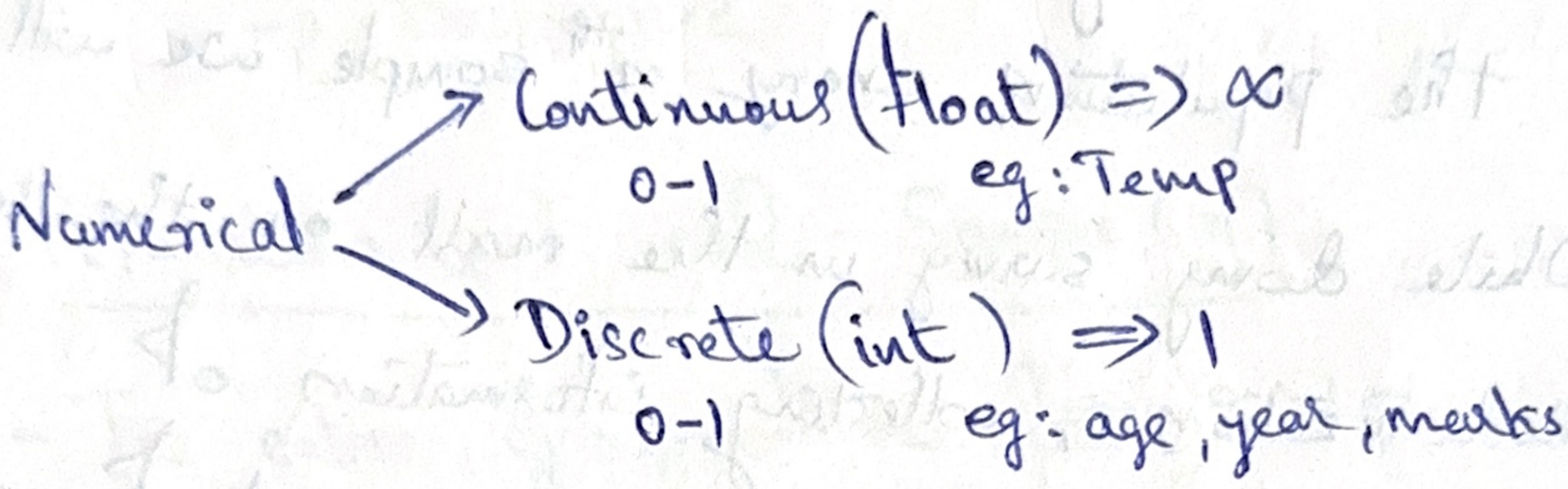
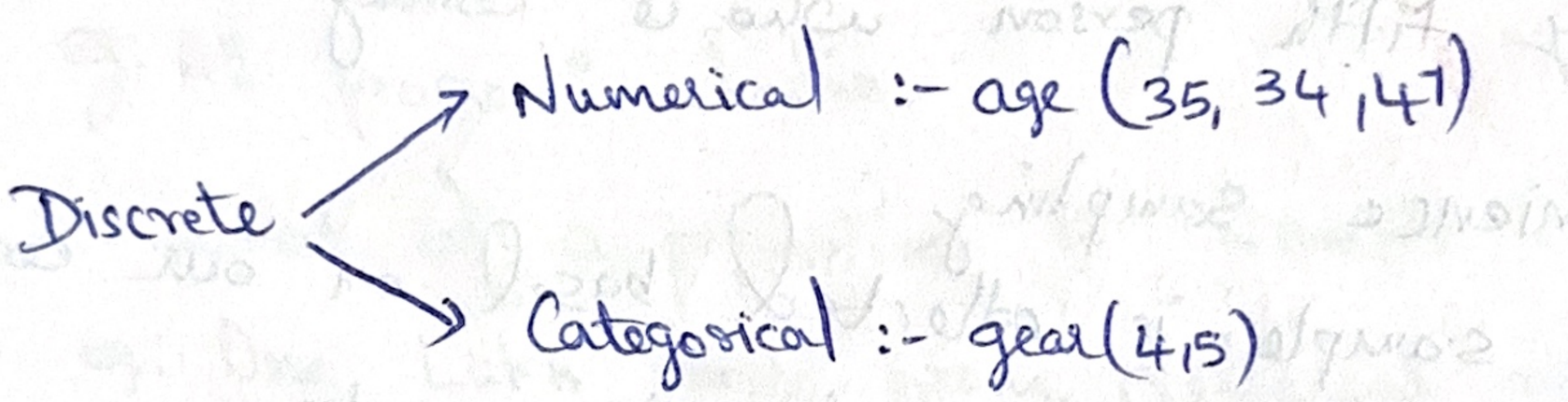
1. Quantitative Variable

A value can be measured & we can perform mathematical operations like (A, S, M, D)

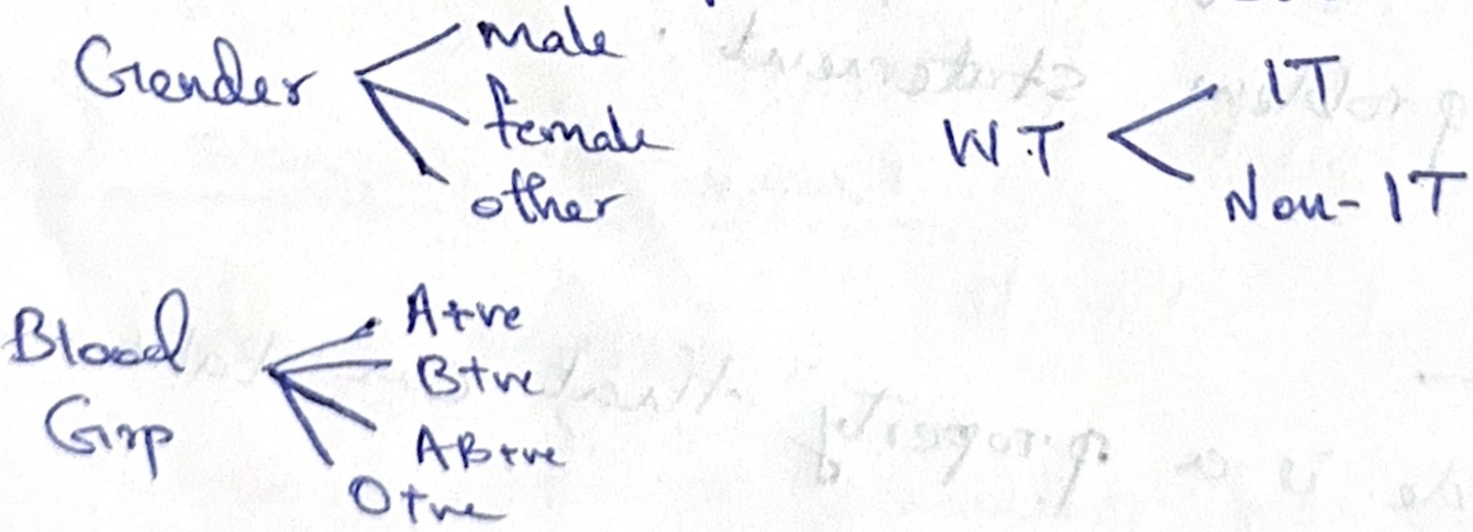
eg: mpg, weight, height





2. Qualitative variable

Non-measurable data and based on some characteristics, we can derive categorical variables.



Variable Measurement Scales

⇒ 4 types of measured variables.

1) Nominal Data:

The categorical data which one having different classes.

2) Ordinal Data:

Order of the data matters but values doesnot

$\{\underset{4}{64}, \underset{3}{73}, \underset{1}{97}, \underset{5}{50}, \underset{2}{82}\}$

$\{\underset{5}{33}, \underset{2}{95}, \underset{4}{55}, \underset{1}{97}, \underset{3}{88}\}$

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.

Descriptive Statistics

1) Measure of central tendancy

Mean Median Mode

Population Mean (μ)

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

Sample Mean ($\bar{X}$)

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Median

* Sort the value either asc or desc order
* choose the mid value
* If you get Mid 2 value, take avg of those 2 value

[1,2,2,3,4,5]

Mean = 2.8

Median = 2.5

[1,2,2,3,4,5,100] → outlier [extreme]

Mean = 16.7

Median = 3

Mean will be affected by outlier whereas median won't affect by outlier

* for null value imputation [purpose]
* ~~Mode: most repetative values~~

Mode

Most repetetive values