__/__/____

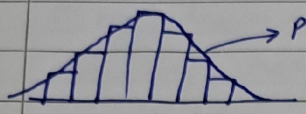
## Statistics

↳ collection, organisation, interpretation, presentation
↳ Decission Making

Age: {1, 2, 3, 4, ...} ⟶ { Distribution, Mean, Median of Age (Analysis of data) }

**Presentation**

① Histogram ② Pdf ③ Cdf




MAIN BRANCH —Statistical Analysis AND Decission→ SUB BRANCH

Application:- ① Everyone uses it ② DATA ANALYSIS
③ MACHINE LEARNING ④ RISK ANALYSIS

E.G:- Vaccination (final) came to market after doing statistical decission.

Statistics

Descriptive

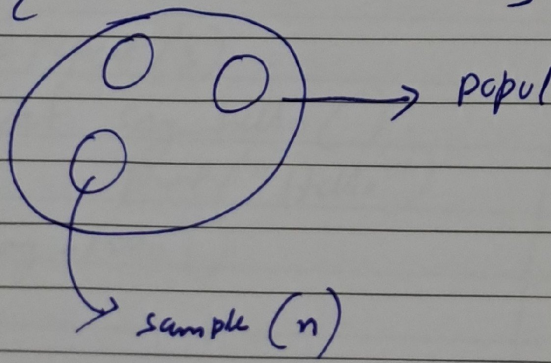
① organizing and summary of data
② Measure of central Tendency
(i) Mean (ii) median (iii) MODE
③ Measure of Dispersion
① Variance ② Standard Deviation

Inferential

① collect data (sample data) { using some experiment } → conclusions or inferences
other data ← { Population data }

E.g:- There is a college A ⟶ (population data) 1000 students (Avg height)

Sample { 100, 200, 300, ... }



## Measure of central tendency

① mean ($\mu$)

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

② Median $\{1, 2, 3, 4, 100\}$ ← outlier (Big impact)

odd → central element

A: $\{1, 2, (3), 4, 100\}$

$= 3$

even → Avg of 2 central element

A: $\{1, 2, (3, 4), 5, 100\}$

$\frac{3+4}{2} = 3.5$

③ Mode (o/p) Max frequency element

A: $\{(4), 3, 2, 1, 1, (4), (4), 5, 2, 100\}$ o/p = 4

## Measure of dispersion — spread ∝ Dispersion

① variance

$$\sigma^2 = \sum_{i=1}^{N} \frac{(x_i - \mu)^2}{N}$$

sample mean

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})}{n-1}$$

in sample variance its $n-1$ (Bessel's correction)

not $n$, coz, if $n \longrightarrow \bar{x} \ll \mu$ $s^2 \ll \sigma^2$

Degree of freedom is given by $n-1$

sample standard deviation = $\sqrt{\text{sample variance}}$

Variable → (Take any value)

- Quantative variable
  - Discrete Quantative (whole No.)
  - Continous Quantative (any Real No.)
- Qualitative / categorical
  - E.g:- Gender → Male, female
  - Colors- Red, Green, Blue

__/__/____

## Random variable (X)

X → functions → values → Process on Experiments

$X = \begin{cases} 0 & H \\ 1 & T \end{cases}$ Tossing a coin

$X = \begin{cases} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{cases}$ Rolling a fair dice

→ Discrete Random (Tossing a coin)

→ continuous Random (Rain data)

① Percentile = $\frac{\text{No. of values below } x}{n} \times 100$

Value = $\frac{\text{Percentile}}{100} \times (n+1)$

② Quantiles 25% = 1st Quantile 50% = 2nd Quantile

75% = 3rd Quantile

IQR → Inter Quantile Range

① ~~MAX~~ MIN ② 1st Quantile ③ Median ④ 3rd Quantile ⑤ MAX

Lower fences $Q1 - 1.5(IQR)$ ⟷ Higher fences $Q3 - 1.5(IQR)$

## Covariance and correlation

① $cov(x,y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n-1}$

$x_i$ → Datapoint of random variable x

$\bar{x}$ → sample mean of n

$y_i$ → Datapoints of random variable y

$\bar{y}$ → Sample Mean of y

② Correlation

Pearson correlation coefficient

$$\rho_{x,y} = \frac{cov(x,y)}{\sigma_x \cdot \sigma_y}$$

Spearman Rank correlation

$$r_s = \frac{cov(R(x), R(y))}{\sigma(R(x)) * \sigma(R(y))}$$