Statistics

① Descriptive

ⓐ Aggregate

max(), sum()...

ⓑ Central Tendency

Mean, Median, Mode

ⓒ Dispersion / Spreadness

Range, IQR, Std

ⓕ Relationship

covariance(), correlation()

- Scatter

ⓓ Data Distribution

Skewness, kurtosis

ⓔ Shape visual

Hist, Boxplot, Violin

- Normal
- Abnormal (skewed)

Probability

① $P(A) = \frac{\#FE}{\#TE}$

② Additive Rule : 1 experiment
2 Events ?

ME : $P(A \text{ or } B) = P(A) + P(B)$

NME : $P(A \text{ or } B) = P(A) + P(B) - P(A \cap B)$

③ Multiplicative : More than 1 exp

same ↖

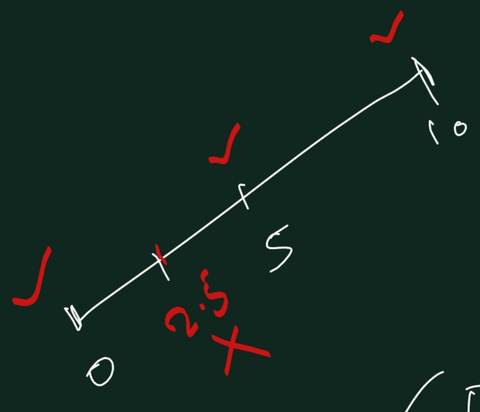
IDE : $P(A \text{ and } B) = P(A) * P(B)$

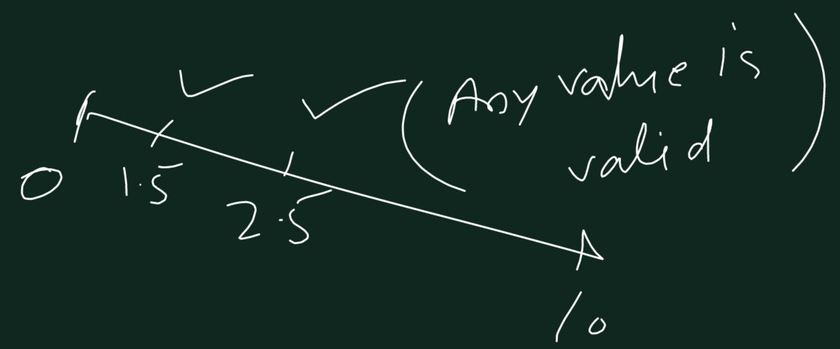
DE : $P(A \text{ and } B) = P(A) * \boxed{P(B/A)}$

↳ reduce

Conditional Probability

Probability Distribution

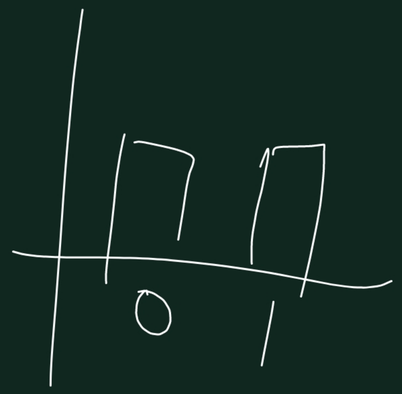




Discrete

(Fixed Values / Text)

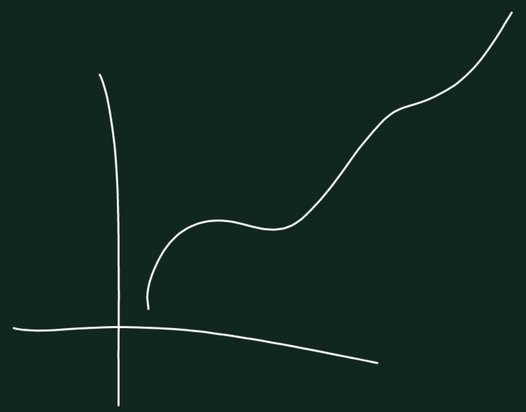
Category



Continuous

(Decimal / Any value)

Numeric



① Discrete

- Bernoulli — 1 trial
- Binomial — Multiple trial
- Poisson — With respect to time

② Continuous

- ✓ Normal (Gaussian, Mean = Median = Mode)
- Uniform
- Exponential
- log
- ✓ Pareto (80 : 20)

Probability Dist Function

1. PMF – Probability Mass Function ⟶ Discrete (x=0, 1)
2. PDF – " Density " ⟶ Continuous (x = 0.3, ---)
3. CDF – Cumulative " ⟶ Both (x ≤ 3)

Permutation & Combination

↓ Ordered

↓ Unordered & Unique

Inferential Statistics

→ inferring/concluding about entire population using a sample.

→ Population , Sample

80:20
70:30
75:25

Sampling Technique

① Convenience → First N / Last N → .head() / slicing [:]

② Systematic → Every nth → slicing [::]

③ Stratified → selection from group

④ Random → Possibility of any record. → random()

(Population)

Case A

Name	Age
A	20
B	21
C	22
D	19
E	20
F	22
G	23

Mean Age (μ) = 21

Std = 1.3 (Low)

my data is normal

Case B (Sample)

A 20

E 20

Mean (μ_s) = 20

Std (σ_s) = 0 (very low)

I can say average age of entire population is $\simeq$ 20.

→ Data is Normal.

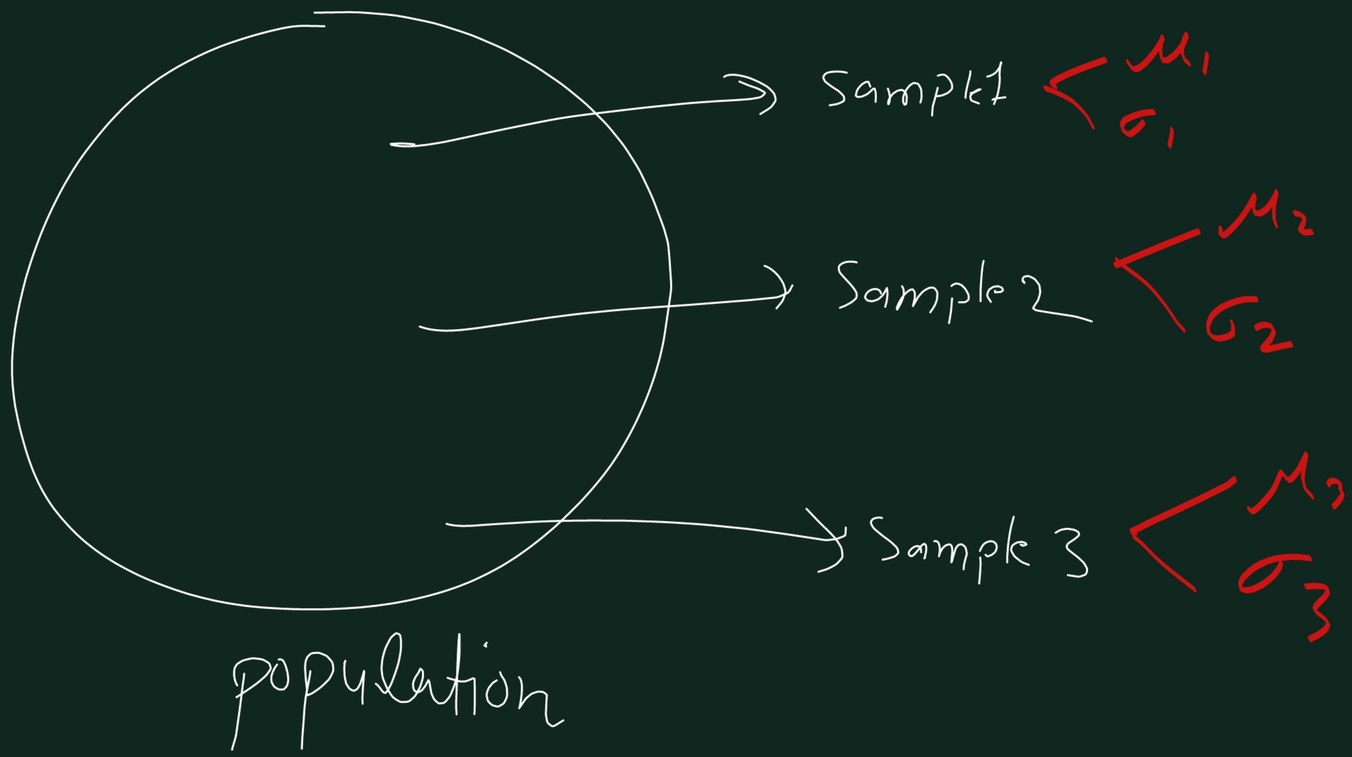
① Let's not take single sample,

↳ Go for multiple sample.

↳ Let's average out

CLT

Central limit theorem



μ_1, σ_1

μ_2, σ_2

μ_3, σ_3

$$\mu_s = \frac{\mu_1 + \mu_2 + \mu_3}{3}$$

$$\sigma_s = \frac{\sigma_1 + \sigma_2 + \sigma_3}{3}$$

#To improve accuracy.