Inferential Statistics

Inference → Giving / providing ideas based on some observation.

Population & Sample:- Population is the universe of data of a particular quantity / variable.

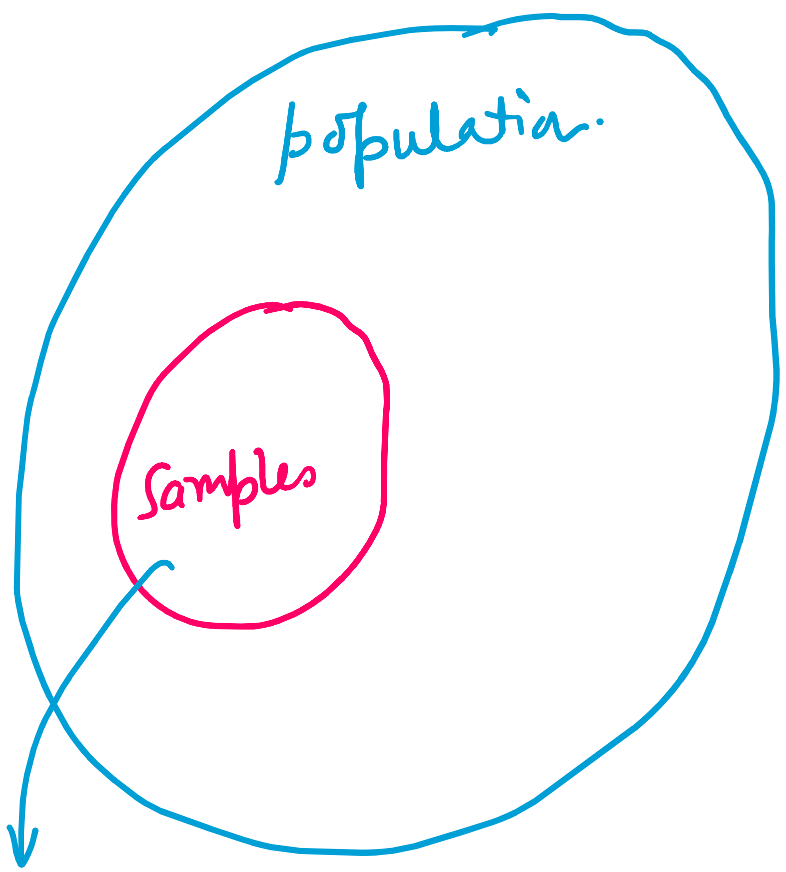
For example:- (1) Suppose we wish to know the average height of adult indian male.

Population:- Heights of all the adult indian male.

(2) Suppose we wish to know how a certain drug XYZ works on a certain disease. i.e. we wish know how much time on average does it take to cure that disease by that drug.

Population:- All the patients records who are having that disease & administered that drug.

Inferential Statistics gives you a way to estimate certain parameter based on the samples you collect.



time ✓
cost ✓
availability ✓

Sample is small fraction / subset of the population whose data we can gather using surveying techniques.

100 different places.

Gather heights of 20 adult males.

2000 datapoints. is a sample of entire adult male population

Population	**Sample**
Parameters	Statistics
Population mean (μ)	Sample mean ($\bar{x}$)
Population Std. deviation (σ)	Sample std. deviation (s)

Sample mean provides an estimate of population mean.

Sample Std. deviation provides an estimate of population Std. deviation

$\bar{x}$ is an estimate of μ

s is an estimate of σ

How to calculate $\bar{x}$ & S :-

$(n > 1)$

Suppose there are 'n' observations in our sample. (Sample size $= n$)

$x_1, x_2, x_3, \ldots\ldots, x_n$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}$$

Sample std. dev. :-

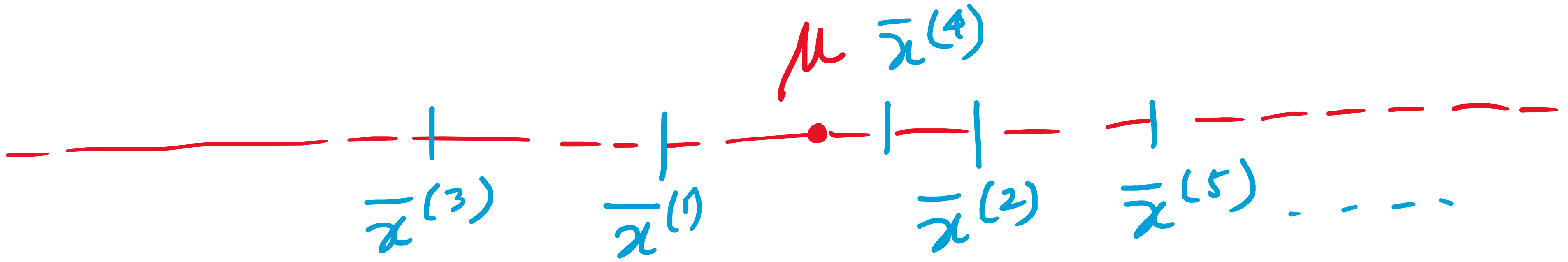
$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Sample variance :-

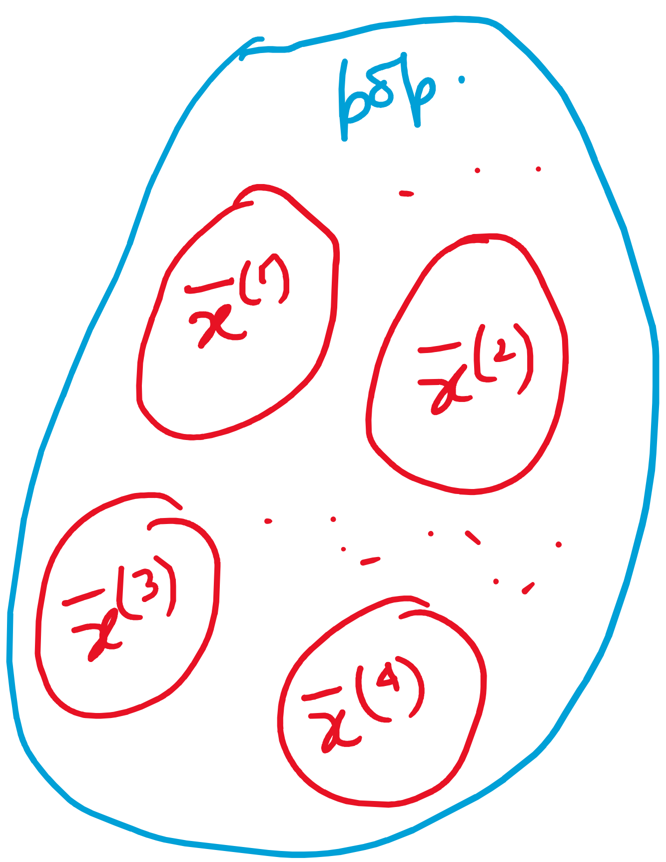
$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

In case of pop. std. deviation we divide by 'n' instead of 'n-1'.

Sample & Population mean:-



If I take different samples from the same population we shall get different sample statistics. & those will be lying around the true population mean μ.



$$\frac{\bar{x}^{(1)} + \bar{x}^{(2)} + \bar{x}^{(3)} + \cdots + \bar{x}^{(m)}}{m}$$

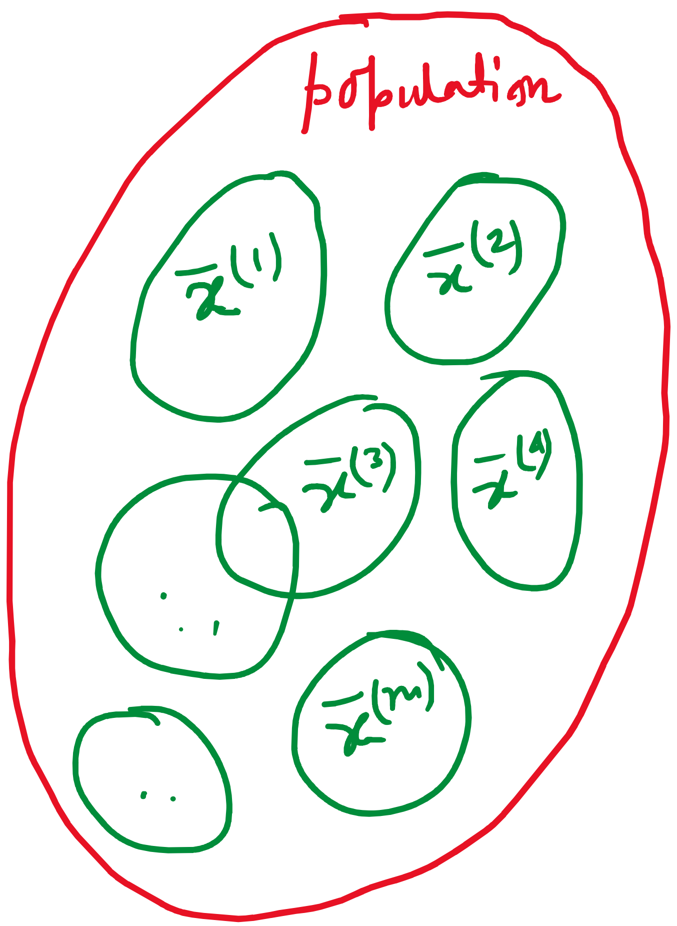
This gives very good estimate of population mean μ.

$$\mu = \mathbb{E}[\bar{x}]$$

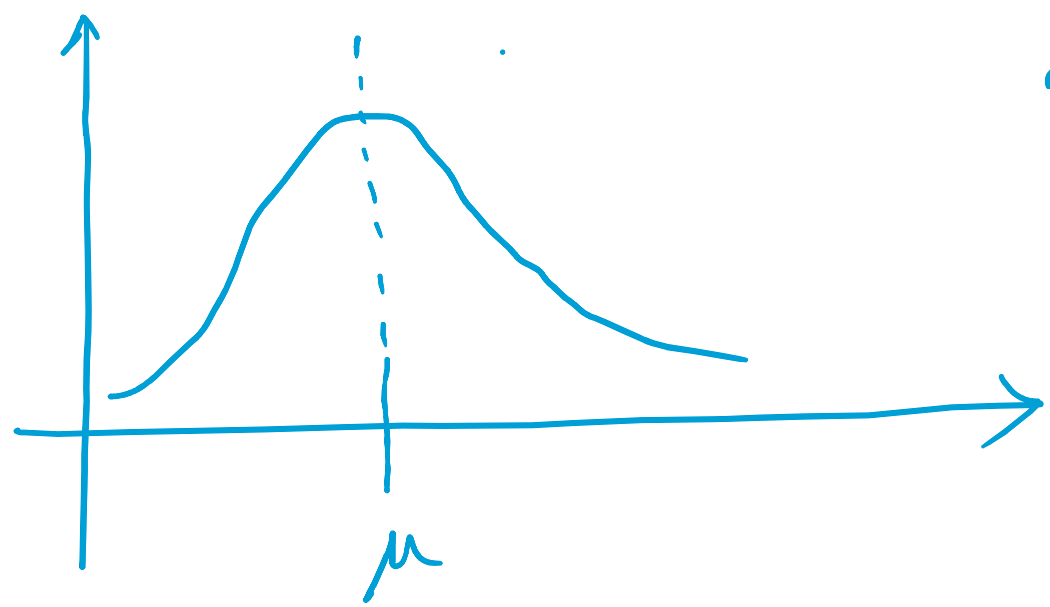
$\bar{x}$ is called <u>unbiased estimator</u> of μ.

Sampling Distribution:-

Sampling Distribution is also known as distribution of sample means.



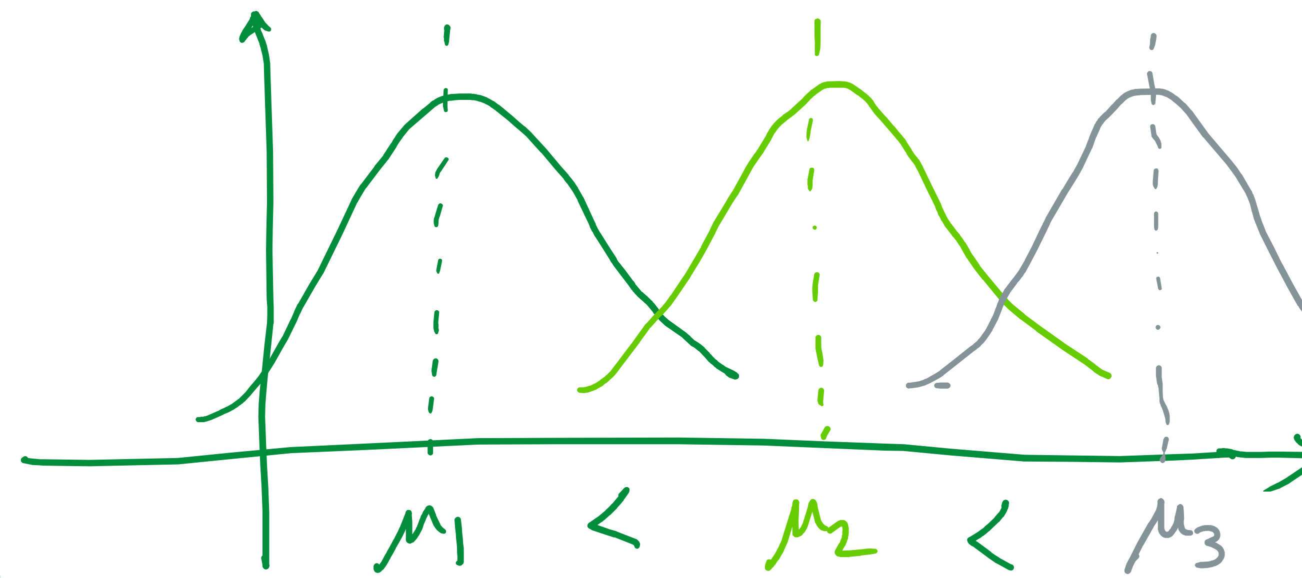
$\left\{ \bar{x}^{(1)}, \bar{x}^{(2)}, \bar{x}^{(3)}, \ldots, \bar{x}^{(n)} \right\}$ $\longrightarrow$ set of sample means obtained from different samples taken from same population.



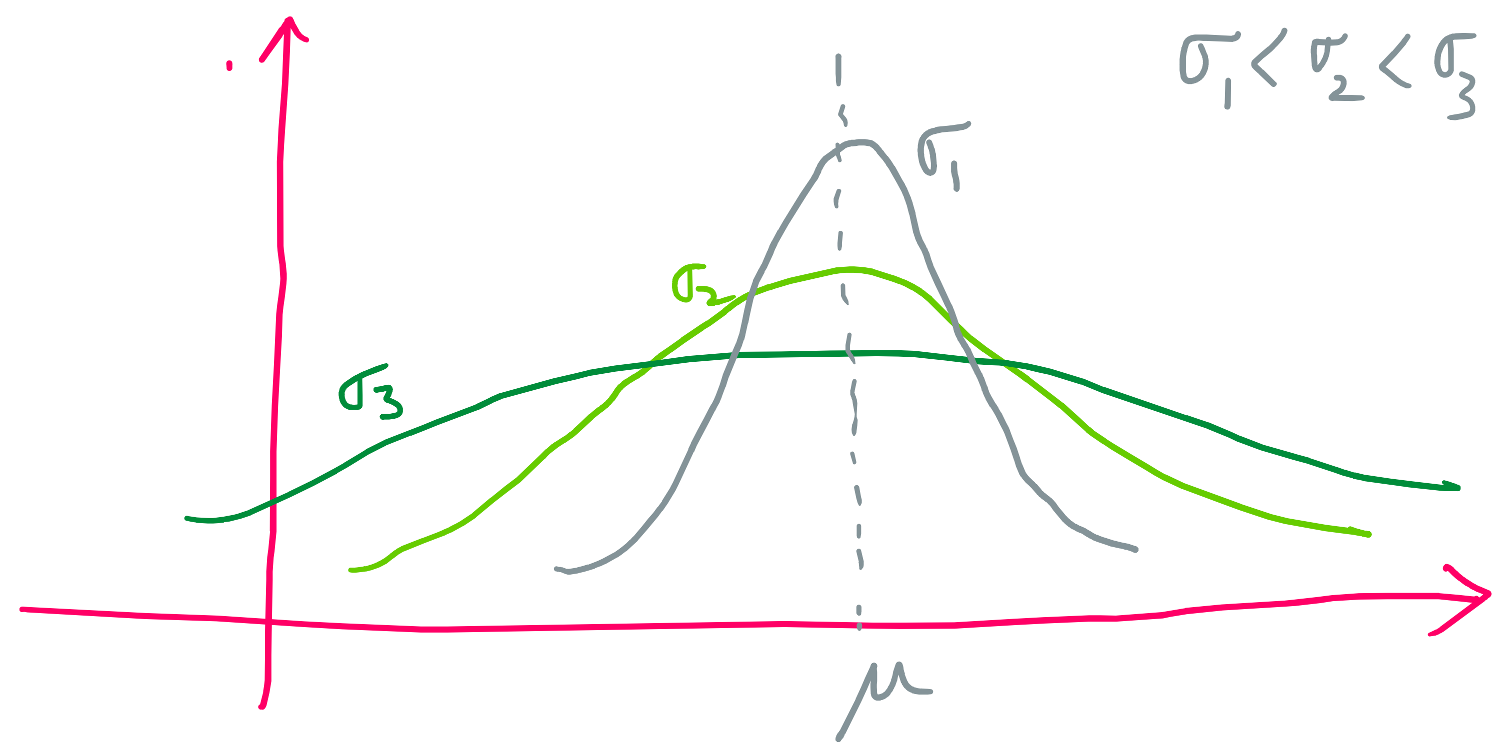
If I plot we will see distribution of sample mean around the population mean.

PDF of Normal distribution:-

$$f(x) = \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}}_{\text{Symmetric around mean.}}$$



$$f(x, \underbrace{\mu, \sigma}_{\text{distribution parameters.}})$$



Higher the 'σ' more flatter will be the curve

Central Limit theorem:-

The sampling distribution or distribution of sample means follow a normal distribution with mean = population mean (μ) & std. dev = $\frac{\text{pop std. dev}}{\sqrt{\text{sample size}}}$.

X is the Random variable which denotes sampling distribution

$$X \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Sampling distribution is nothing but a normal distribution.

with mean = population mean & standard dev = $\frac{\text{population std. dev}}{\sqrt{\text{sample size}}}$

Standard error.

If the sample size increases ↑ → standard error ↓

sample size (n)	μ	SE ($\sigma/\sqrt{n}$) ($\sigma = 10$)
10	60	$10/\sqrt{10} = \sqrt{10} = 3.16$ (SE_1)
25	60	$10/\sqrt{25} = \frac{10}{5} = 2$ (SE_2)
36	60	$10/\sqrt{36} = \frac{10}{6} = 1.6$ (SE_3)
100	60	$10/\sqrt{100} = \frac{10}{10} = 1$ (SE_4)