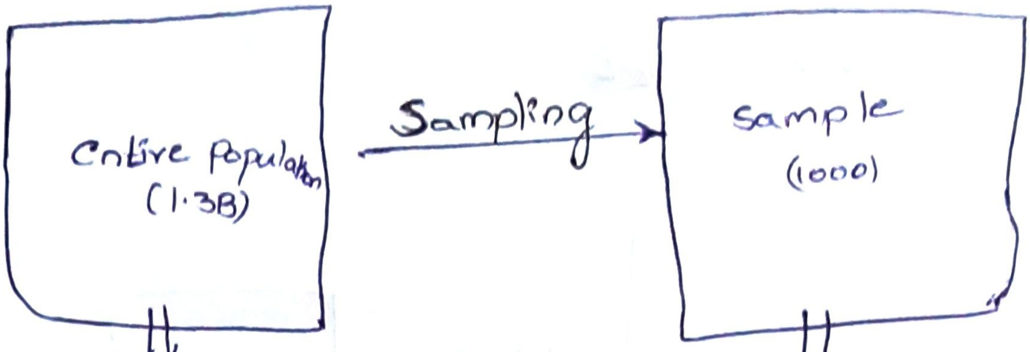
Inferential statistics:-



mean $= \mu$	mean $= \bar{x}$
Variance $= \sigma^2$	Variance $= S^2$
Std. Deviation $= \sigma$	Std. deviation $= S$

$\boxed{\mu \approx \bar{x}}$

Sampling technique

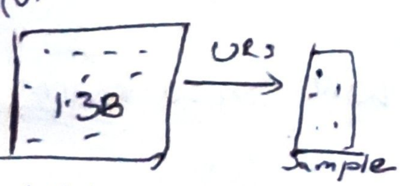
If sample is biased then it is not correct (i.e. $\mu \neq \bar{x}$)

Sampling technique

If sample is not biased; then it is correct (i.e $\mu = \bar{x}$)

Sampling techniques:-

1. Convinience Sampling :- → biased
2. Volunteer Sampling → biased.
3) Uniform Random Sampling (URS) → un biased

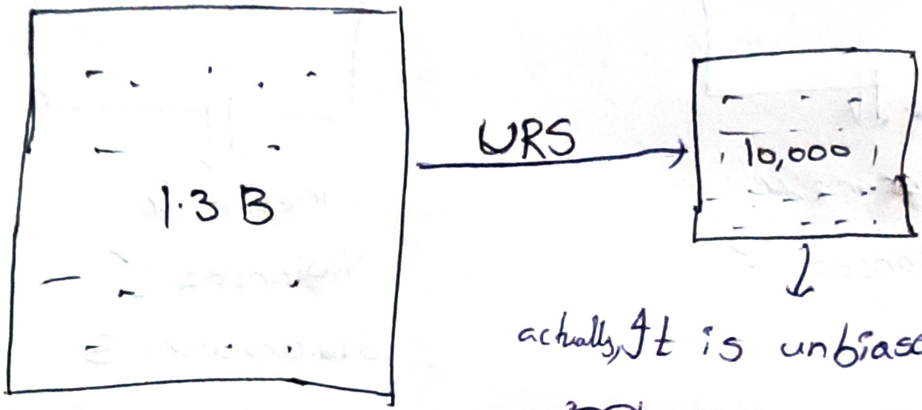


you can pick the any people from the entire population, but the probability should be same.

$\boxed{\mu \approx \bar{x}}$

point estimate

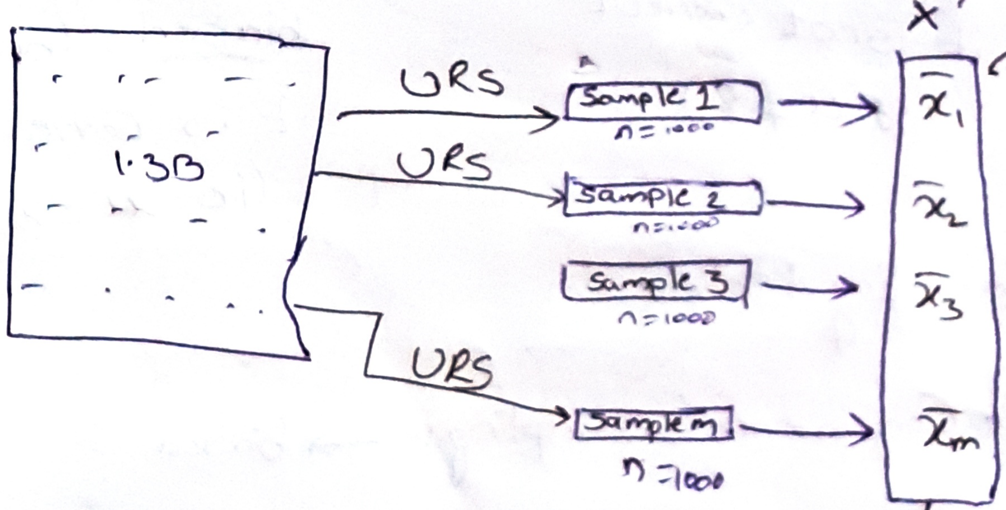
Its a point



actually, It is unbiased; If we have outlier in sample then it becomes biased.

To avoid outliers we are splitting; i.e

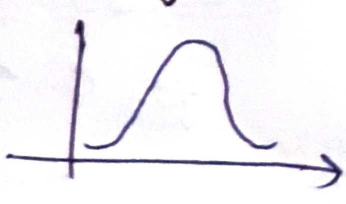
$\bar{X}$ → sns.distplot



Sampling Distribution (or) Distribution of sample mean.

always the "sampling Distribution" will show "normal distribution"

plot of sampling Distribution



$$\boxed{\mu_{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \bar{x}_3 + \cdots + \bar{x}_m}{m}}$$

mean of sampling distribution.

<u>observations:-</u>

1. Sampling Distribution will form a Normal Distribution.

$$\underbrace{\bar{x}}_{\text{Sampling Distribution}} \overset{\text{follows}}{\sim} \underbrace{N}_{\text{normal}}\left(\overset{\text{mean}}{\mu_{\bar{x}}}, \underbrace{\frac{\sigma^2}{n}}_{\text{variance}}\right)$$

2.) Mean of Sampling Distribution $\rightarrow \mu_{\bar{x}}$

$$\mu \approx \mu_{\bar{x}}$$

3.) std deviation of Sampling distribution $\rightarrow \frac{\sigma}{\sqrt{m}}$

where $\sigma \Rightarrow$ population S.D

Population: N

Sample: n $\rightarrow$ $\bar{x} = \frac{\Sigma(\text{obs})}{n}$

$$s^2 = \frac{\Sigma(\text{obs} - \bar{x})^2}{n-1}$$

$$\mu = \frac{\Sigma(\text{obs})}{N}$$

$n \rightarrow$ no of samples in sample

$N \rightarrow$ no of samples in population

variance $\left\{ \sigma^2 = \frac{\Sigma(\text{obs} - \mu)^2}{N} \right.$

1) point estimate:

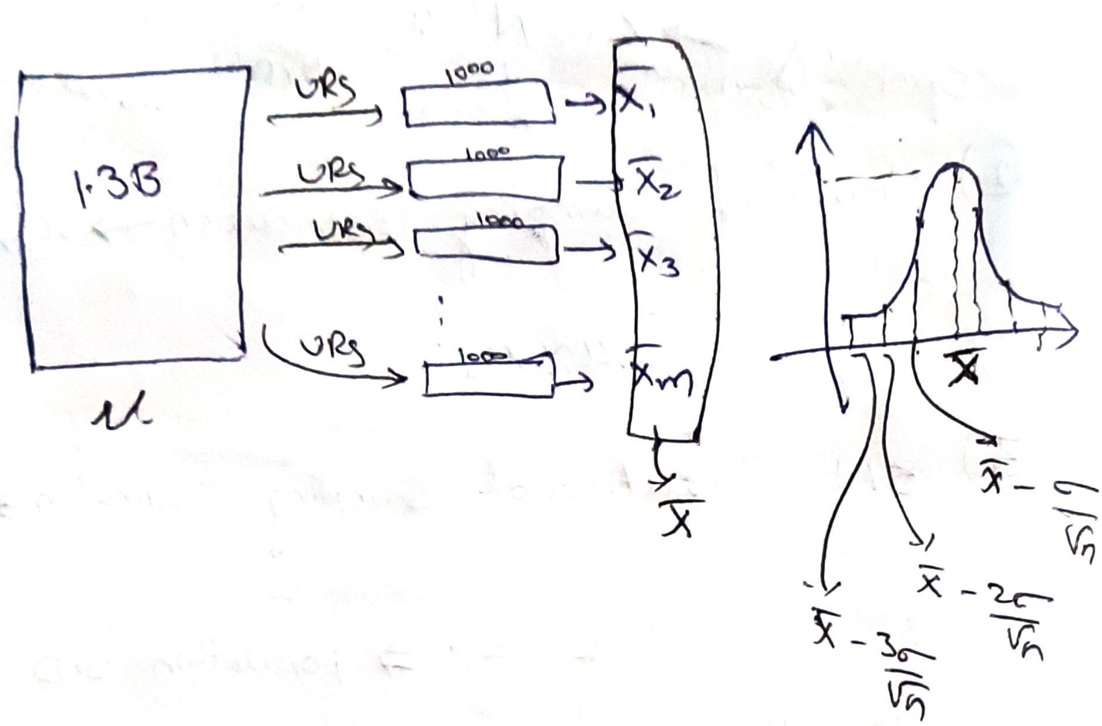
$\cancel{\bar{x}}$

$\mu \approx \bar{x}$

2) Interval estimate:

$\mu = [\bar{x} - error, \bar{x} + error]$

with y% Confidence level



$\mu = [\bar{x} - error, \bar{x} + error]$ with y% of confidence level

$\mu = [\bar{x} - \frac{\sigma}{\sqrt{n}}, \bar{x} + \frac{\sigma}{\sqrt{n}}]$ with 68% of confidence level

$\mu = [\bar{x} - \frac{2\sigma}{\sqrt{n}}, \bar{x} + \frac{2\sigma}{\sqrt{n}}]$ with 95% of confidence level

$\mu = [\bar{x} - \frac{3\sigma}{\sqrt{n}}, \bar{x} + \frac{3\sigma}{\sqrt{n}}]$ with 99.7% of Confidence level.

$\therefore$ $\boxed{\mu \approx \left[\bar{x} \pm \left(z^* \frac{\sigma}{\sqrt{n}}\right)\right] \text{ with } y\% \text{ Confidence}}$

z^* → Z-score $\Rightarrow$ $z^* = 1 \Rightarrow 68\%$

$z^* = 2 \Rightarrow 95\%$

$z^* = 3 \Rightarrow 99.7\%$

→ If $n < 30$; we have to use

n → Samplesize

"<u>Students-t distribution</u>" (not Central Limit Thm)

⇓

$$\mu \approx \left[\bar{x} \pm t_{n-1,\alpha/2} \frac{\sigma}{\sqrt{n}}\right] \text{ with } y\% \text{ Confidence}$$

σ → S.D of population

(but $\boxed{\sigma \approx S}$ → Sample S.D)

σ → Population S.D

$t_{n-1,\alpha/2}$ → critical value.

$n-1$ → Degrees of freedom

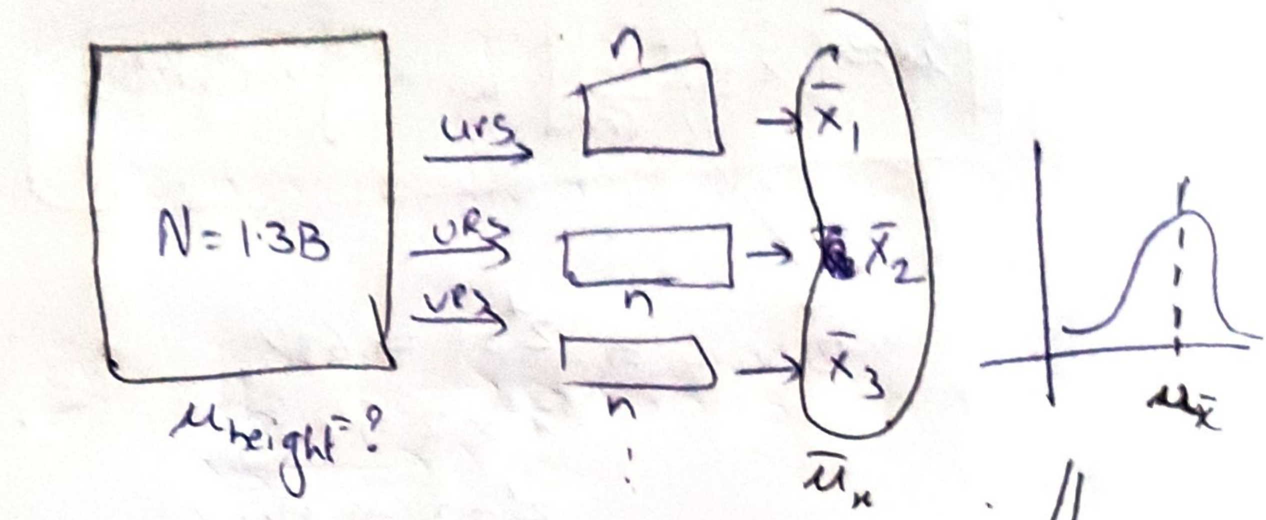
<u>Ex:-</u> If we have 5 samples (samples not in population)

with 95% confidence

$\Rightarrow$ 95% $\Rightarrow P = 0.95 \Rightarrow 1 - \alpha = 0.95 \Rightarrow \alpha = 0.05$

$\alpha/2 = 0.025$

$n = 5$ samples // $t_{n-1,\alpha/2} = t_{4,0.025}$



Point estimate ← $\boxed{\mu_{\bar{x}} \approx \mu_{height}}$

Interval estimate ~~Z-score~~ ← $\boxed{\mu_{height} \approx [\mu_{\bar{x}} \pm Z_{score} \text{ std error}]}$

with 95% confidence

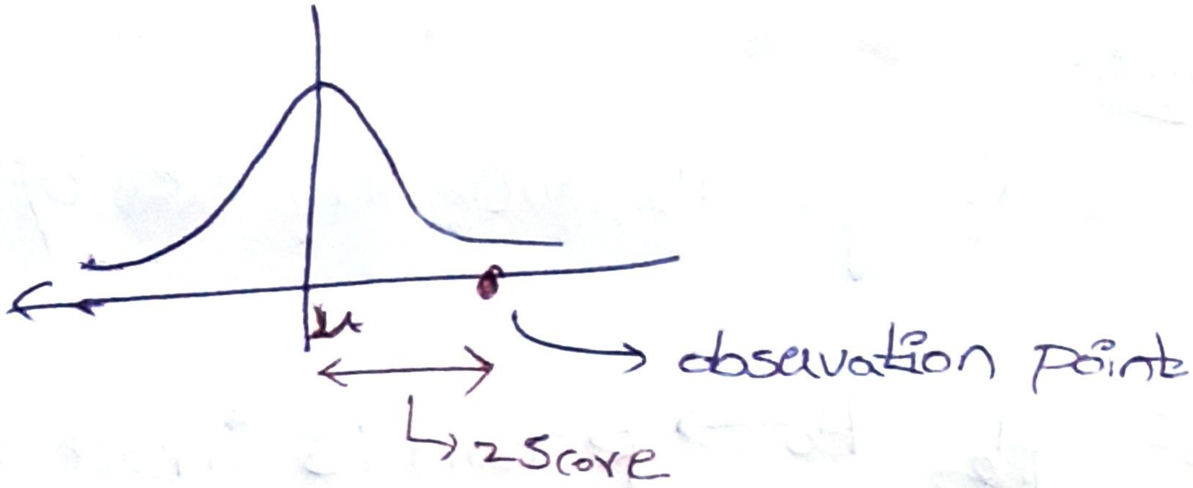
If size of sample < 30 then use t-score

else use Z-score

t-Score

↳ $\mu_{height} \approx [\mu_{\bar{x}} \pm t_{n-1,\alpha/2} \text{ std error}]$

Z-score:- How far away the observation from mean with respect to std



Note:- ① "t-distribution" is also known as

"students t-distribution"

$t_{(n-1), \alpha/2}$ → degrees of freedom → critical value.

② we use t-distribution, when

i) $n < 30$ ⇒ n ⇒ sample size

(or)

(ii) σ is not known.