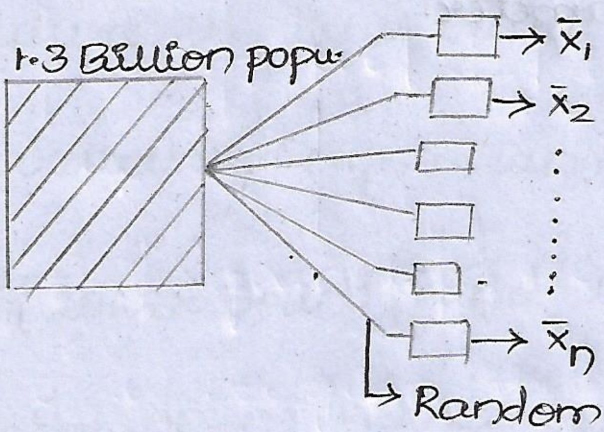
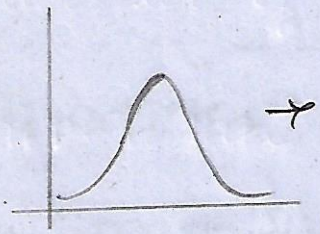
LECTURE - 35



(fairly large means it need to be greater than 100)



→ Then it is called as a sampling distribution ⇒ $\bar{x} \sim N[\mu_{\bar{x}}, std]$

where $\mu_{\bar{x}}$ - sampling distribution of mean.

std - sampling distribution of std.

$$\sigma^2 = \frac{\sum_{i=1}^{N=1.3B} (\mu - \text{datapoint})^2}{1.3 \text{ Billion}}$$ → Average squared distance from the mean.

variance (σ^2) → measure of spread from the mean.

's' is considered as std, then

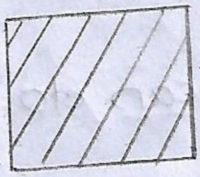
$$s^2 = \frac{\sum_{i=1}^{n} (\bar{x}_1 - \text{datapoint})}{n-1}$$

NOTE: n-1 is done for variance.

The above (s^2) sampling is taken since it is biased. we

thats the reason we will take 'n-1'.

CLT makes a assumption :

 $\rightarrow \mu = ?, \sigma = \_\_\_$

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

2. If we know the spread of the data, then

   CLT is used with the μ & σ

3. So, $\bar{x} \sim N\left(\mu_{\bar{x}}, \frac{\sigma}{\sqrt{n}}\right)$

For the above assumptions,

↳ The sampling distribution shows a bell shape curve.

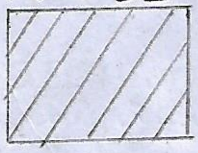
↳ Mean of sampling distribution μ_x is going to be equivalent to populations mean.

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

↳ std of sampling distribution is $\frac{\sigma}{\sqrt{n}}$ where 'n' is the size of the sample.

QUESTION :

N = 1.3B



$\rightarrow \mu \in [\_,\_]$ with CI = ?

$\mu_n = ?$

If σ is given, then we can do CLT

If σ is not given, then we can do CLT but it takes lot of time.

STEP-1: Take a random sample from the entire population.

STEP-2: Calculate the mean of the sample.

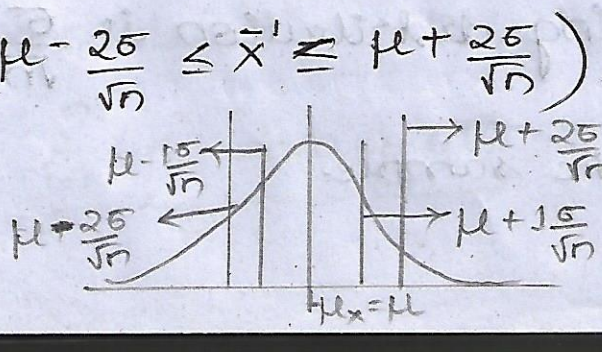
$$\bar{x}' = \frac{\Sigma \text{ datapoint}}{\text{Total no. of datapoints } (n)}$$

$\bar{x}' \rightarrow$ single sample.

STEP-3: $\mu \approx \bar{x}$ - point estimate.

STEP-4: If we know the std of the population, then we can apply CLT

$$P\left(\mu - \frac{2\sigma}{\sqrt{n}} \le \bar{x}' \le \mu + \frac{2\sigma}{\sqrt{n}}\right) = 95\%$$



$$\Rightarrow P\left(\bar{x}' - \frac{2\sigma}{\sqrt{n}} \leq \mu \leq \bar{x}' + \frac{2\sigma}{\sqrt{n}}\right) = 95\%$$

$$\Rightarrow \mu \in \left[\bar{x}' - \frac{2\sigma}{\sqrt{n}}, \bar{x}' + \frac{2\sigma}{\sqrt{n}}\right]$$ with 95% of confidence.

$$\Rightarrow \mu \in \left[\bar{x}' - \frac{1\sigma}{\sqrt{n}}, \bar{x}' + \frac{1\sigma}{\sqrt{n}}\right]$$ with 68% of confidence.

where $\frac{\sigma}{\sqrt{n}}$ is std of sampling distribution.