Lecture 1

Normal & Chi Distribution

Sampling Distribution

Population $\rightarrow$ μ, σ^2 parameters

Sample $\longrightarrow$ $\bar{x}$, s^2 statistic $\longrightarrow$ what distribution of s^2

What distribution of $\bar{x}$?

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_n \end{bmatrix}_{k \times 1} \qquad \phi = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \\ \phi_n \end{bmatrix}_{k \times 1}$$

Not distribution of population but of sample

C	x
1	x_1
2	
n	x_n

$\left.\begin{matrix} \bar{x} \\ s^2 \end{matrix}\right\}$ Fixed value but distribution comes when repeated multiple times

$\rightarrow$ Statistic a a RV

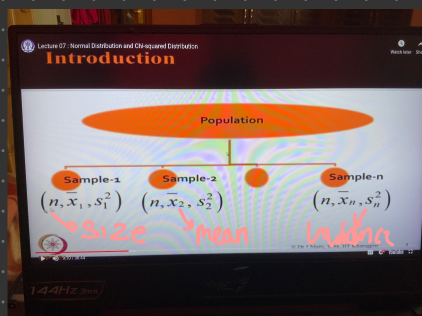
(1)

$\bar{x}_{smallest}$

Let it be $\bar{x}_{(1)}$

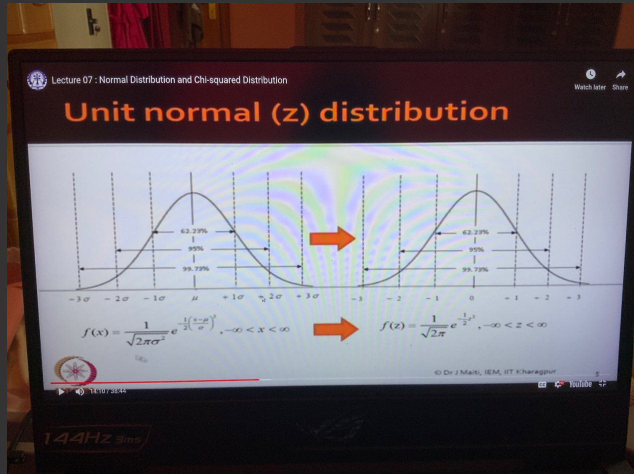
$$\bar{x} = [\bar{x}_1, \bar{x}_2 \quad \bar{x}_n]^T$$

$$\bar{x}_{()} = [\bar{x}_{(1)} \ \bar{x}_{(2)} \quad \bar{x}_{(n)}]$$



$z = \frac{x-\mu}{\sigma}$ [transformed]

↑ Standardised Variable



$$\text{Standardised Variable} = \frac{x - E(x)}{\sqrt{E(x-\mu)^2}}$$

$$\boxed{\text{Mean} = E(z)} = E\left(\frac{x-\mu}{\sigma}\right) = \frac{E(x) - E(\mu)}{\sigma} = \frac{\mu - \mu}{\sigma} = 0$$

$$V(z) = E(z - E(z))^2 = E(z^2) = E\left(\left(\frac{x-\mu}{\sigma}\right)^2\right) = \frac{1}{\sigma^2}\left[E(x-\mu)^2\right]$$

$$\boxed{V(z) = 1}$$

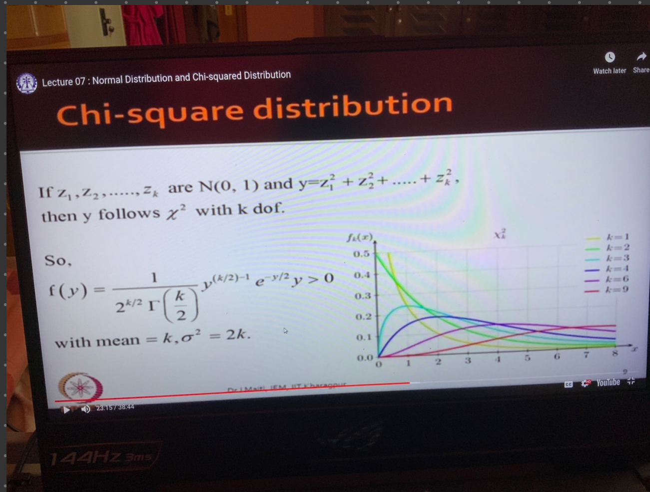
$$= \frac{\sigma^2}{\sigma^2} = 1$$

Look at example 22 24

CHI-SQUARE DISTRIBUTION

Suppose we know there is a normal variable, its square is taken then summed over and uses for different purpose

↳ That sum follows a distribution called Chi-Square Distribution



z - Not affect by degrees of freedom

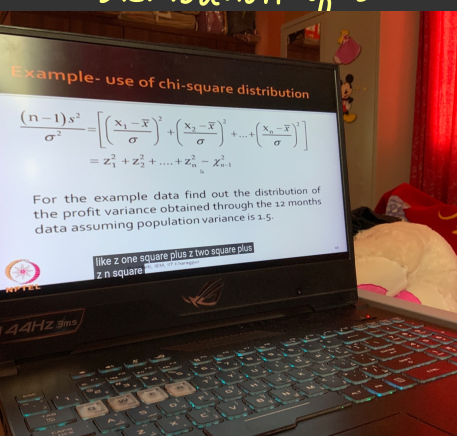
Σz^2 - Affected with k DOF

s^2 = Variance,

Distribution of $\bar{x}$: If x is normally dist, $\bar{x}$ is also Normal

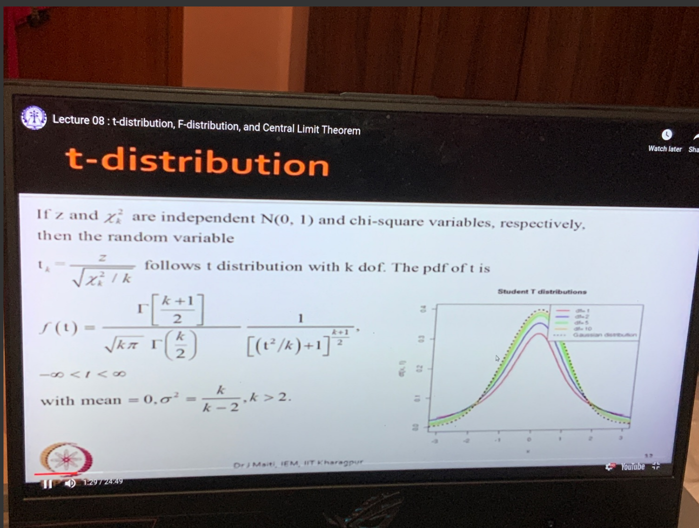
$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

normal variable u manipulate



When we use internal distribution we use chi-square distribution

Lecture 2



2 variables → independent

↳ chi-square

$$t_k = \frac{z}{\sqrt{\chi_k^2 / k}}$$

↳ follow t-distn with k degrees of freedom

When DOF = ∞, t-distribution = normal

Use:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

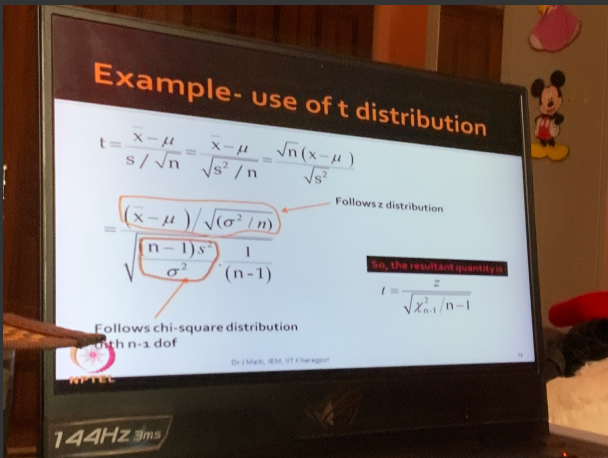
$$E(\bar{x}) = E\left[\frac{1}{n}\sum x_i\right] \quad \frac{n\mu}{n} = \mu$$

$$Var(\bar{x}) = V\left[\frac{1}{n}\sum x_i\right] = \frac{1}{n^2}\left[n\sigma^2\right] = \frac{\sigma^2}{n}$$

$$z = \frac{\bar{x} - E(\bar{x})}{Var(\bar{x})} = \frac{\bar{x} - \mu}{\sqrt{\frac{\sigma^2}{n}}} - \frac{\bar{x} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

Sample std deviation



Follows chi-square distribution

F-Distributions:

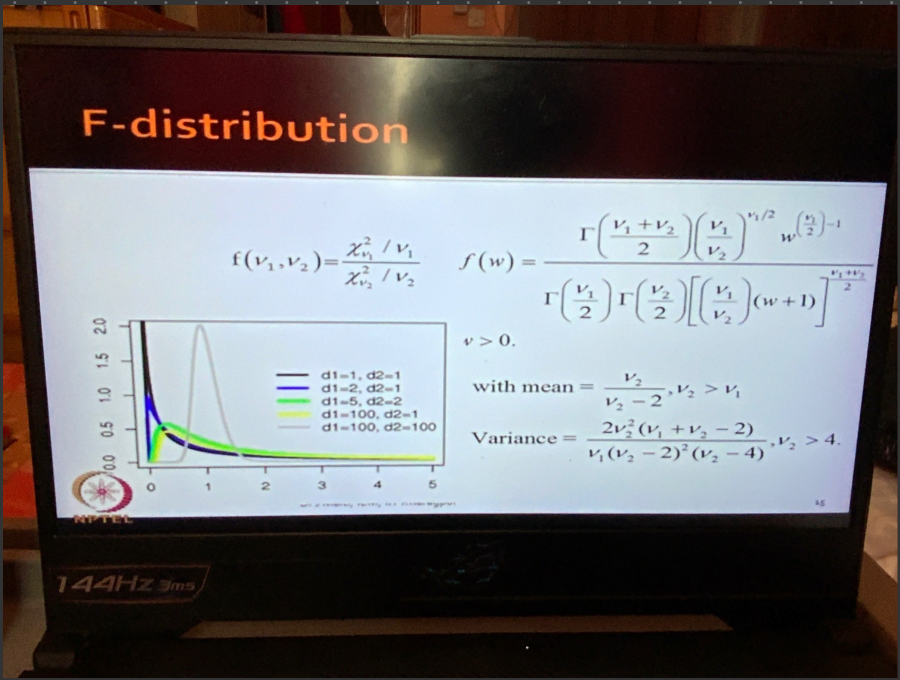
If there a 2 distributions describing a population P_1, P_2 b s/a me Varible

$$F = \frac{S_1^2}{S_2^2}$$

S_1^2 ← Pop P_1

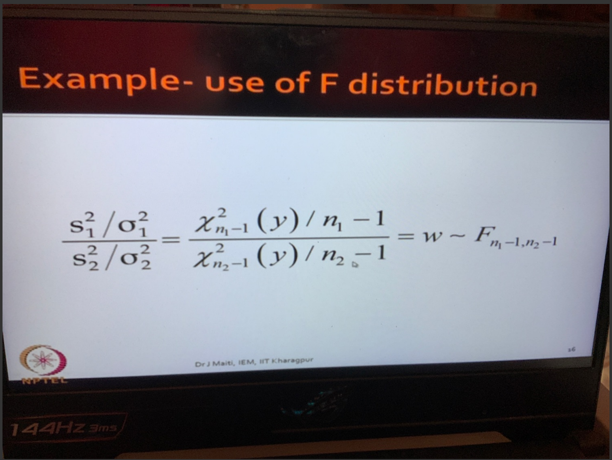
S_2^2 ← Pop P_2

We assume at times $s_1^2 = s_2$



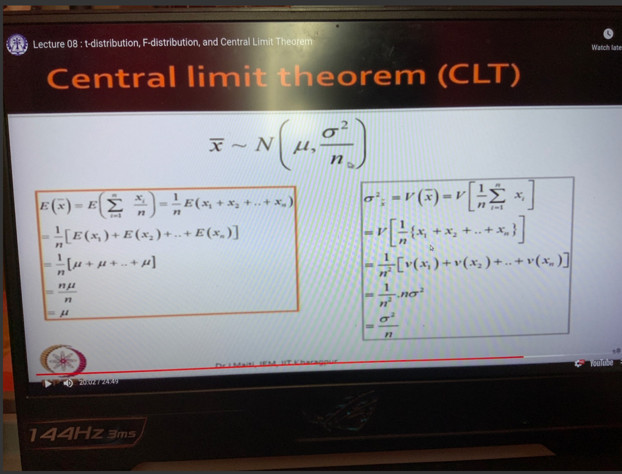
Characterised by 2 degrees of freedom variables d_1, d_2

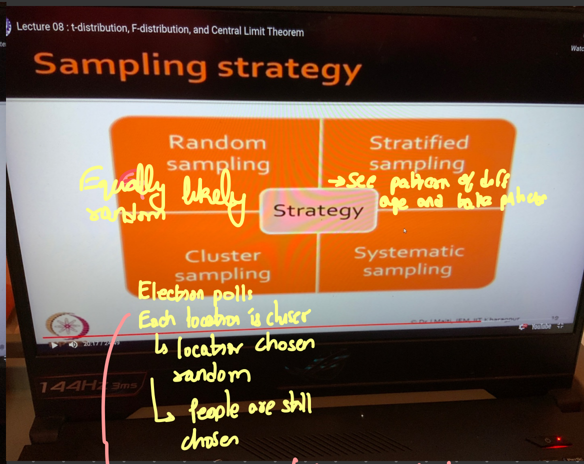
→ Num DOF, Den DOF



Central Limit Theorem

→ If we sample from normal population, then sampling distribution of $\bar{x}$ is normal





Equally likely random

→ see pattern of diff age and take pictures

Election polls
Each location is cluster
↳ location chosen random
↳ people are still chosen

How was data collected

$\boxed{\text{Watch 21-24:0}}$

Lecture 3

Estimation [under univariate statistics]

- → Point
  - → Mean
  - → SD [Variance]
- → Interval
  - → Single population
    - → CI for mean
    - → CI for variance
  - → Two population
    - → CI for diff of mean
    - → CI for ratio of two possible variance

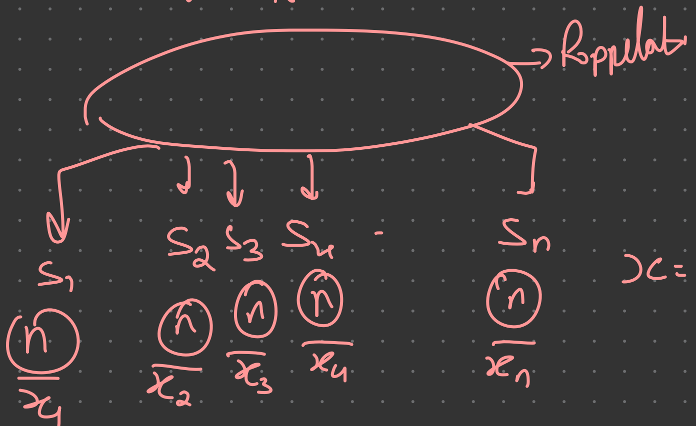
$$x = \begin{bmatrix} x_1 \\ x_2 \\ | \\ x_n \end{bmatrix} \qquad \bar{x} = \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Estimation of population mean

Point estimate of Mean

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

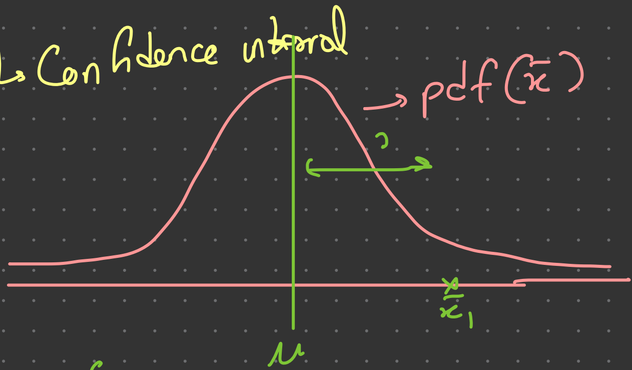
Point estimate of variance



$$x = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \\ \bar{x}_n \end{bmatrix} \rightarrow \text{Certain distribution}$$

Confidence in estimate:

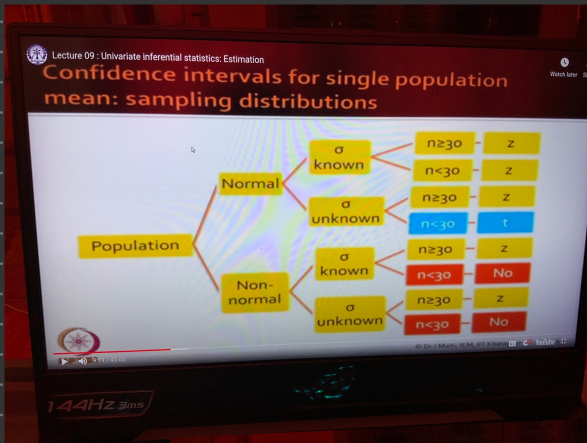
↳ Confidence interval



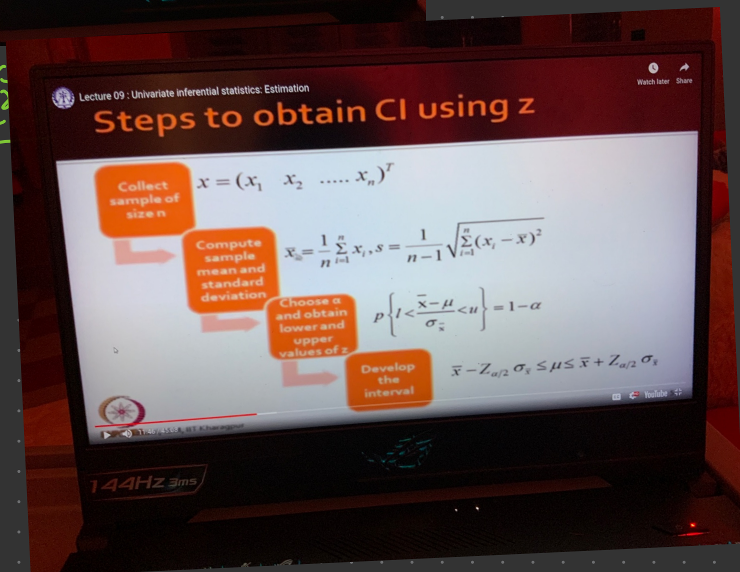
$E(\bar{x}) = \mu$

To know if $\bar{x}_1$, is representative of μ, it depends on SD from μ

To know if interval presents, we look at distribution



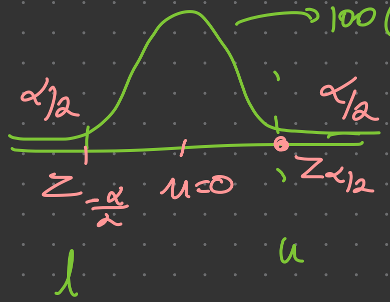
$$Z = \frac{\bar{x} - E(\bar{x})}{\sigma_{\bar{x}}}$$



$$P\left\{ l < \frac{\bar{x} - E(\bar{x})}{\sigma_{\bar{x}}} < u \right\} = 1 - \alpha$$

$$P\left\{ l < z < u \right\} = 1 - \alpha$$

$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$



↳ Significance level

How to get l, u?

If we have α, then $\alpha/2$ is known

$$-Z_{\alpha/2} \le \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \le Z_{\alpha/2}$$

$$\Rightarrow -Z_{\alpha/2}\left[\frac{\sigma}{\sqrt{n}}\right] \le \bar{x}-\mu \le Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow -\bar{x} - Z_{\alpha/2}\left[\frac{\sigma}{\sqrt{n}}\right] \le -\mu \le -\bar{x} + Z_{\alpha/2}\left[\frac{\sigma}{\sqrt{n}}\right]$$

$$\bar{x} - Z_{\alpha/2}\left[\frac{\sigma}{\sqrt{n}}\right] \le \mu \le \bar{x} + Z_{\alpha/2}\left[\frac{\sigma}{\sqrt{n}}\right]$$

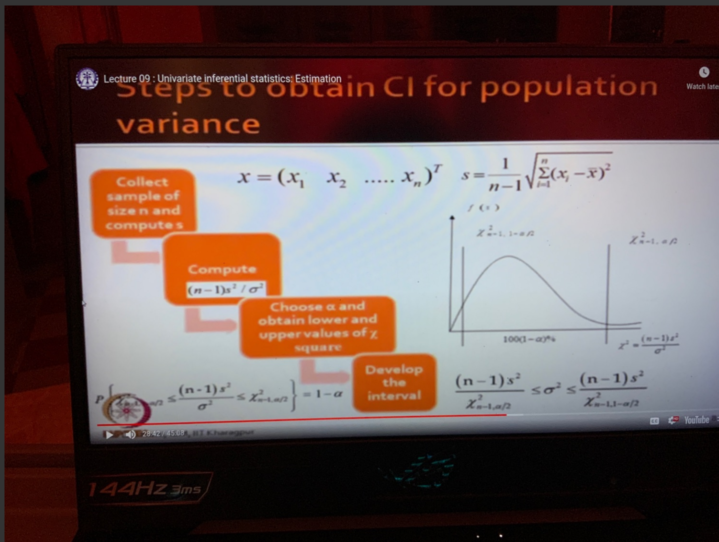
Confidence interval
$100(1-\alpha)\%$

NOTE:
When we say CI, its about population parameters & not sample statistics

If we are using t-distribution,

$$t_{n-1} = \frac{\bar{x}-\mu}{s/\sqrt{n}}$$

$$\bar{x} - t_{n-1,\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_{n-1}(\alpha/2)\frac{s}{\sqrt{n}}$$



Confidence Interval for

$$P(l \le \sigma^2 \le u) = 1-\alpha$$

CHI-SQUARE

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

pdf $f(\chi^2)$ - $\frac{(n-1)s^2}{\sigma^2}$

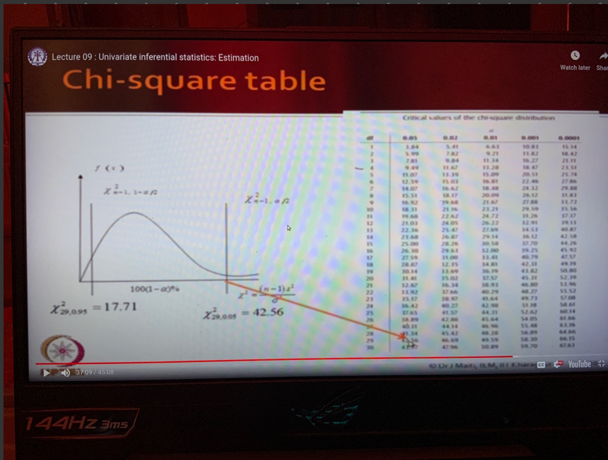
100(1-α)%

$-\frac{\alpha}{2} = l$ $\quad u \quad$ $u = \alpha/2$

$\chi^2_{1-\alpha/2,n-1}$ $\quad$ $\chi^2_{\alpha/2,n-1}$

$$\chi^2_{1-\frac{\alpha}{2}, n-1} \le \frac{(n-1)s^2}{\sigma^2} \le \chi^2_{\frac{\alpha}{2}, n-1}$$

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2},n-1}}$$



Lecture

CI for $\mu_1 - \mu_2$

$$(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \le \mu_1 - \mu_2 \le (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

- σ_1, σ_2 must be known
- n_1, n_2 for normal population

If, $\boxed{\sigma_1 = \sigma_2 = \sigma}$ → population variances are same

⇓

Another estimation

↳ first find $\hat{\sigma} = S_{pooled}$

↳ $\hat{\sigma}_1 = \hat{\sigma}$

↳ $\hat{\sigma}_2 = \hat{\sigma}$

$$S_p = S^2_{pooled} = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2} = \hat{\sigma}^2$$

$$\cancel{\bar{x}_1-\bar{x}_2 - Z_{\alpha/2}\left[s_1\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right]}$$

$$\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} = \sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$
$$= s_1\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

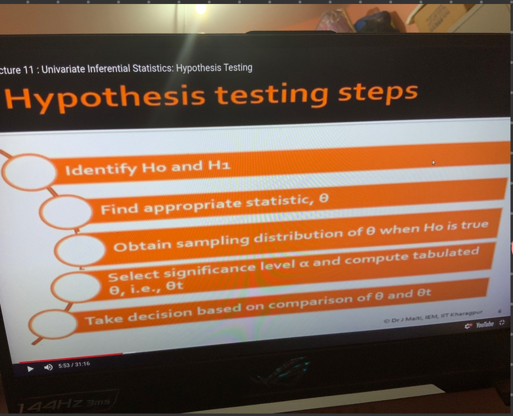
$$(\bar{x}_1-\bar{x}_2) - Z_{\alpha/2}\left[s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right] \le \mu_1-\mu_2 \le (\bar{x}_1-\bar{x}_2) + Z_{\alpha/2}\, s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

Lecture 11

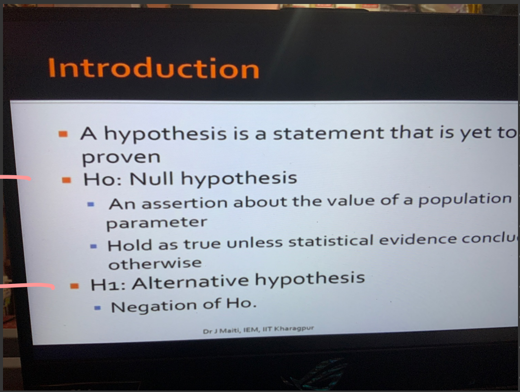
Hypothesis testing:

Statement that is yet to be proven

Population parameters



① ②



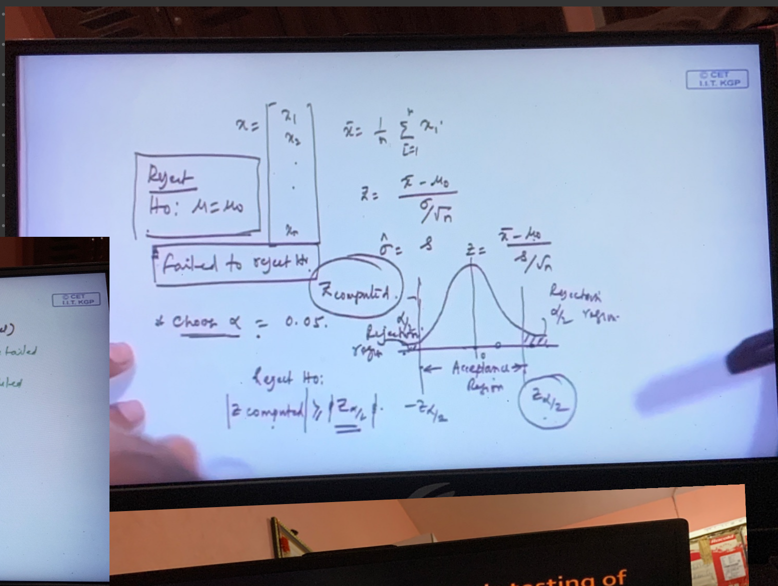
$$z = \frac{\bar{x} - E(\bar{x})}{\sigma/\sqrt{n}}$$

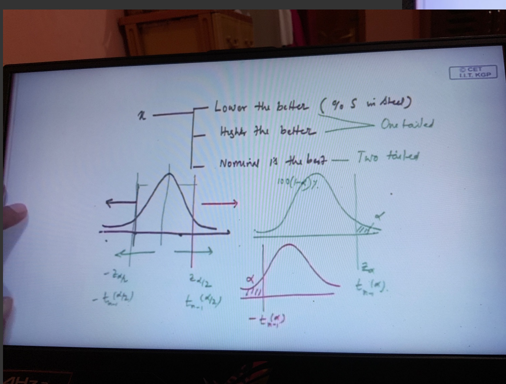
$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

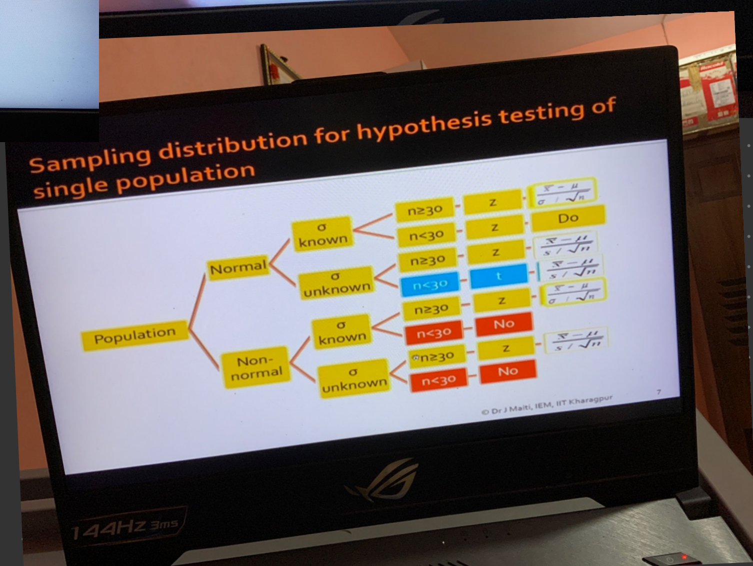
→ Assuming H_0 is true

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{2}}$$

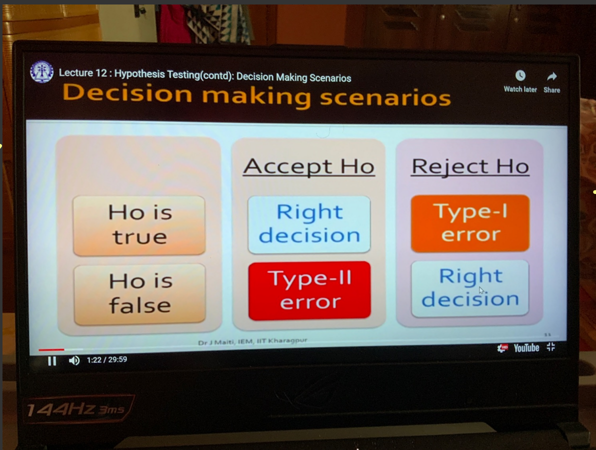
→ Statistic for which we need to know the mean of



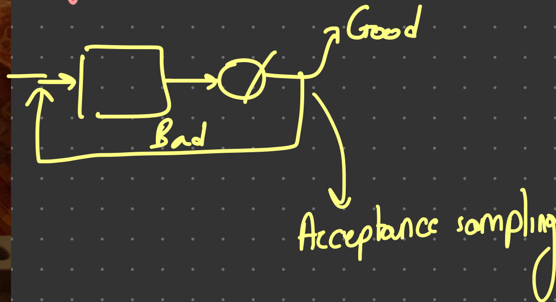




Decision Making Scenarios:



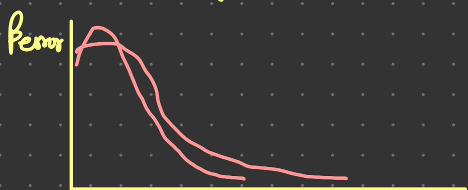
Physical Interpretation:

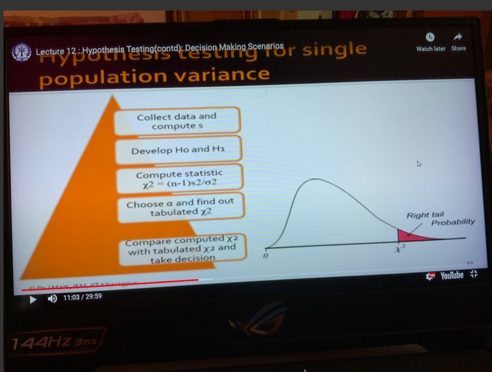


Operating Characteristic Curve [OC Curve]

β $\quad 1-\beta$: Power of test

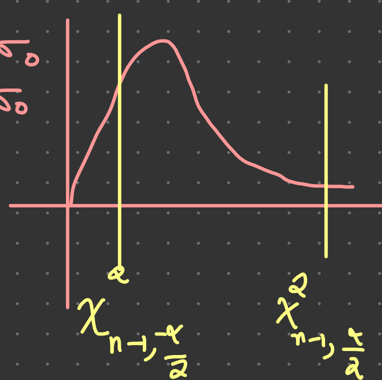
Not able to identify mean diff if too small

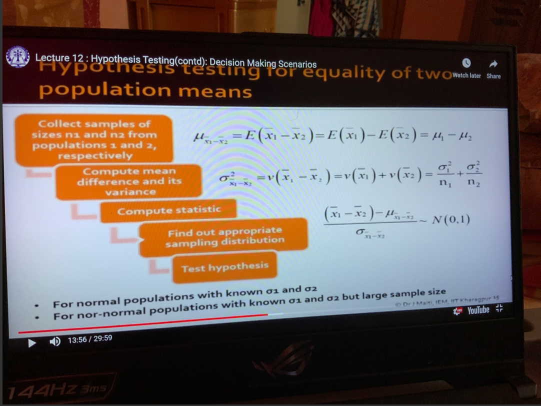




$H_0: \sigma = \sigma_0$

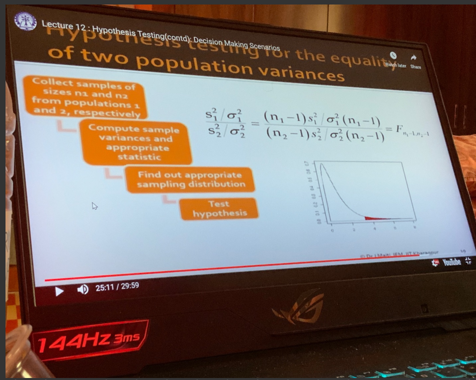
$H_1: \sigma \neq \sigma_0$





$H_0 = \mu_1 = \mu_2$

$H_1 = \mu_1 \neq \mu_2$



$\boxed{\text{Book: Freund \& Miller}}$