

Lecture 1

Normal & Chi Distribution

Sampling Distribution

Population $\rightarrow \mu, \sigma^2$ parameters

Sample $\rightarrow \bar{x}, s^2$ statistic

What distribution of s^2 ?

What distribution of \bar{x} ?

$$\Theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \vdots \\ \theta_n \end{bmatrix}_{n \times 1} \quad \phi = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_n \end{bmatrix}_{n \times 1}$$

Not distribution of population
but of sample

x
 x_1
 x_2

n \bar{x}_n
 s^2 γ fixed value but distribution comes when repeated multiple times

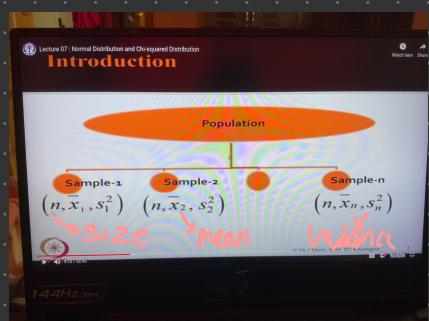
\rightarrow Statistic $\propto \alpha P V$

$\bar{x}_{\text{smallest}}$

debtive $\bar{x}_{(1)}$

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

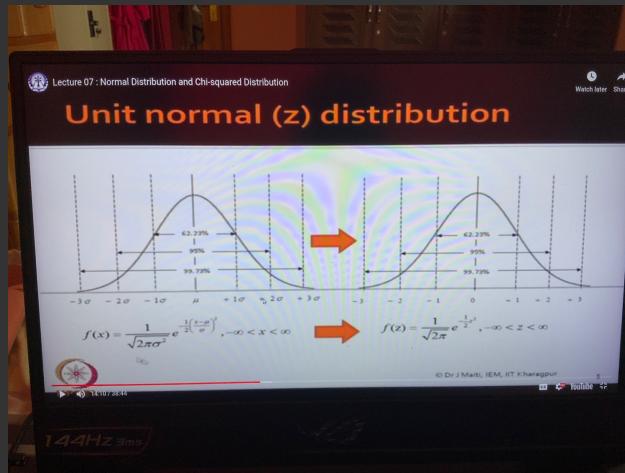
$$\bar{x}_{(1)} = [\bar{x}_{(1)}, \bar{x}_{(2)}, \dots, \bar{x}_{(n)}]$$



$$Z = \frac{x - \mu}{\sigma} \quad [\text{transformed}]$$

↑
Standardised Variable

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



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

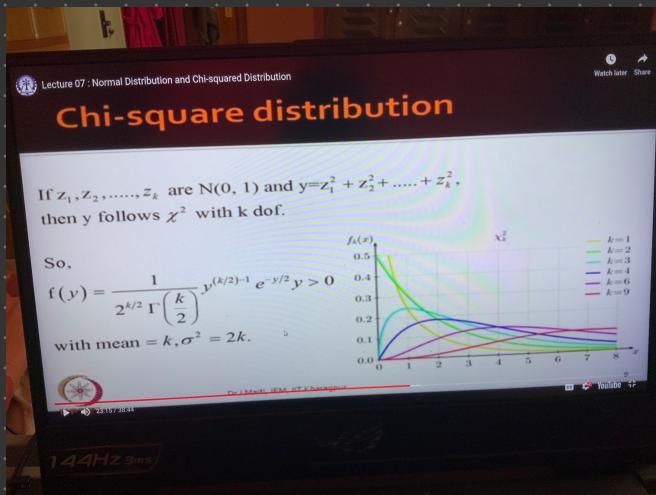
$$\begin{aligned} V(z) &= E(z - E(z))^2 = E(z^2) - E(z)^2 = E\left(\left(\frac{x-\mu}{\sigma}\right)^2\right) = \frac{1}{\sigma^2} \left[E\left((x-\mu)^2\right)\right] \\ &= \frac{\sigma^2}{\sigma^2} = 1 \end{aligned}$$

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



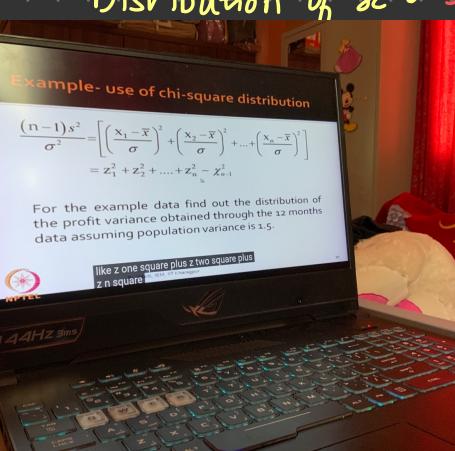
Σz^2 - Not affected by degrees of freedom
 $\sum z^2$ - Affected with k DOF

S^2 = Variance,

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

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

normal variable is manipulated



When we use internal distribution we use chi-square distribution

Lecture 2

Lecture 08 : t-distribution, F-distribution, and Central Limit Theorem

t-distribution

If z and χ_k^2 are independent $N(0, 1)$ and chi-square variables, respectively, then the random variable

$$t_k = \frac{z}{\sqrt{\chi_k^2 / k}}$$
 follows t distribution with k dof. The pdf of t is
$$f(t) = \frac{\Gamma\left[\frac{k+1}{2}\right]}{\sqrt{k\pi} \Gamma\left(\frac{k}{2}\right)} \frac{1}{\left[\left(t^2/k\right)+1\right]^{\frac{k+1}{2}}}$$
$$-\infty < t < \infty$$

with mean = 0, $\sigma^2 = \frac{k}{k-2}$, $k > 2$.

Dr J Maiti IEM, IIT Kharagpur
1297/28/09

2 variables \rightarrow independent
↳ Chi-square

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

follow t-distr with k degrees of freedom

when DOF = ∞ , t-distribution = normal

Use:

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

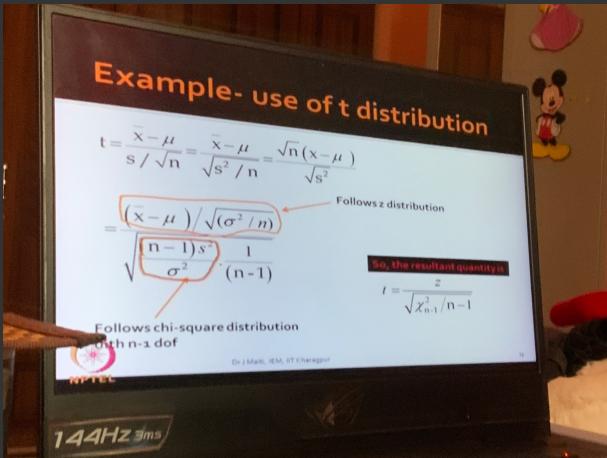
$$E(\bar{x}) = E\left[\frac{1}{n} \sum x_i\right] = \frac{n\mu}{n} = \mu$$

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

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

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

↳ Sample std. deviation



Follows chi-square distribution

F-Distributions:

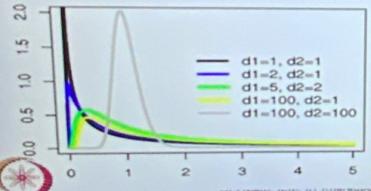
If there are 2 distributions describing a population P_1, P_2 b
same variable

$$F = \frac{s_1^2}{s_2^2} \leftarrow \begin{matrix} \text{Pop } P_1 \\ \text{Pop } P_2 \end{matrix}$$

We assume ab times $s_1^2 = s_2^2$

F-distribution

$$f(\nu_1, \nu_2) = \frac{\chi_{\nu_1}^2 / \nu_1}{\chi_{\nu_2}^2 / \nu_2} \quad f(w) = \frac{\Gamma\left(\frac{\nu_1 + \nu_2}{2}\right)\left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} w^{\left(\frac{\nu_1}{2}\right)-1}}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)\left[\left(\frac{\nu_1}{\nu_2}\right)(w+1)\right]^{\frac{\nu_1+\nu_2}{2}}}$$



$\nu > 0$.

$$\text{with mean} = \frac{\nu_2}{\nu_2 - 2}, \nu_2 > \nu_1$$

$$\text{Variance} = \frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)}, \nu_2 > 4.$$

Characterized by 2 degrees of freedom variables d_1, d_2

→ Num DOF, Den DOF

Example- use of F distribution

$$\frac{s_1^2 / \sigma_1^2}{s_2^2 / \sigma_2^2} = \frac{\chi_{n_1-1}^2(y) / n_1 - 1}{\chi_{n_2-1}^2(y) / n_2 - 1} = w \sim F_{n_1-1, n_2-1}$$

Central Limit Theorem

→ If we sample from normal population, then distribution is normal

sampling of \bar{x}

^
^



Lecture 08 : t-distribution, F-distribution, and Central Limit Theorem

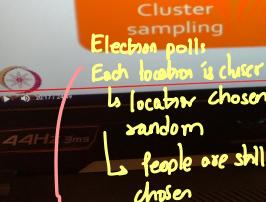
Central limit theorem (CLT)

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\begin{aligned} E(\bar{x}) &= E\left(\sum_{i=1}^n \frac{x_i}{n}\right) = \frac{1}{n} E(x_1 + x_2 + \dots + x_n) \\ &= \frac{1}{n} [E(x_1) + E(x_2) + \dots + E(x_n)] \\ &= \frac{1}{n} [n\mu + \mu + \dots + \mu] \\ &= \frac{n\mu}{n} \\ &= \mu \end{aligned}$$

$$\begin{aligned} \sigma^2_{\bar{x}} &= V(\bar{x}) = V\left[\frac{1}{n} \sum_{i=1}^n x_i\right] \\ &= V\left[\frac{1}{n} \{x_1 + x_2 + \dots + x_n\}\right] \\ &= \frac{1}{n^2} [V(x_1) + V(x_2) + \dots + V(x_n)] \\ &= \frac{1}{n^2} \cdot n\sigma^2 \\ &= \frac{\sigma^2}{n} \end{aligned}$$

Sampling strategy



[Watch 21-24]

Lecture 3

Estimation [Under univariate statistics]



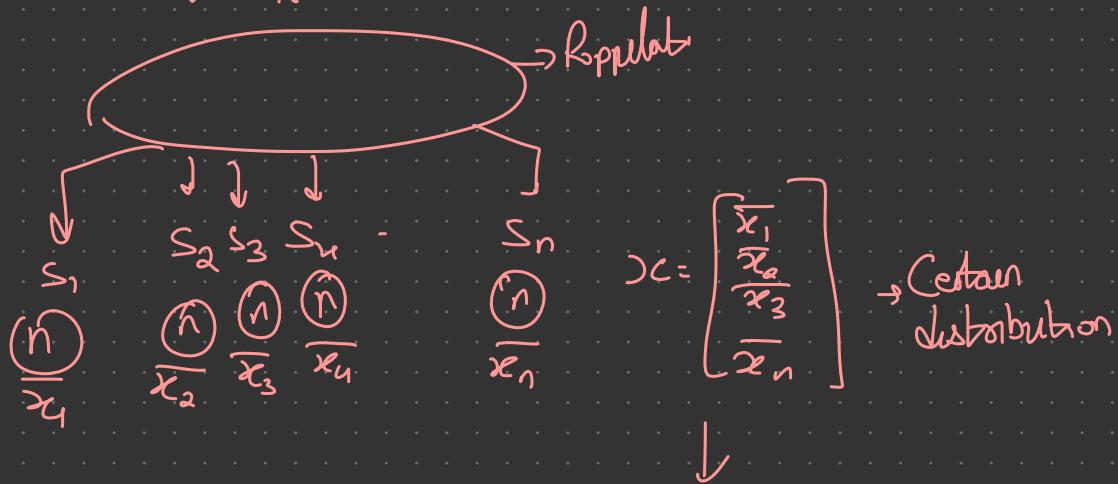
$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \bar{x} - \hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i$$

↑ Estimation of population mean

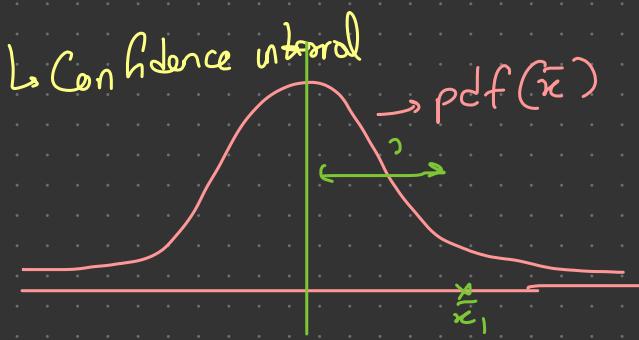
Point estimate of Mean $S^2, \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$

↓

Point estimate of Variance

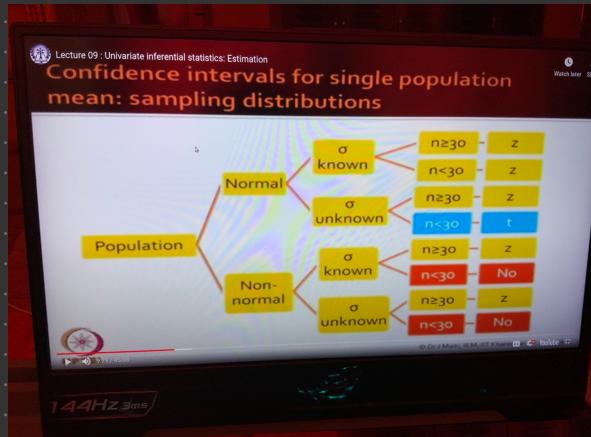


Confidence in estimate:

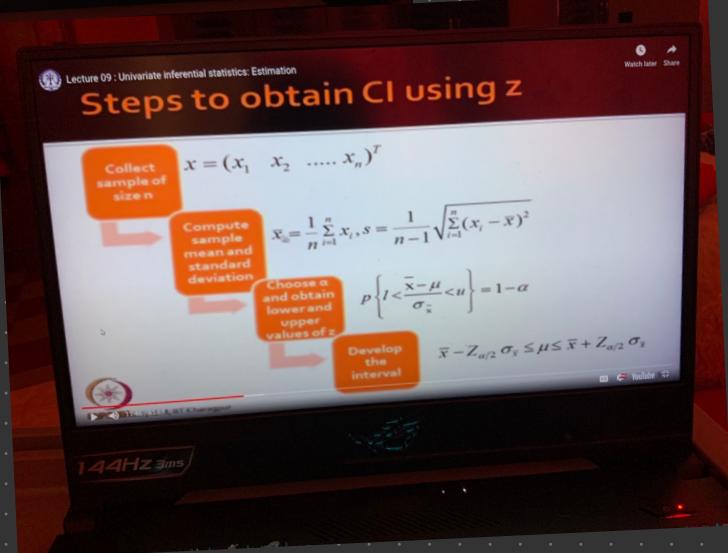


To know if \bar{x}_1 is representative of μ , it depends on SD from μ

To know if interval presents, we look at distribution



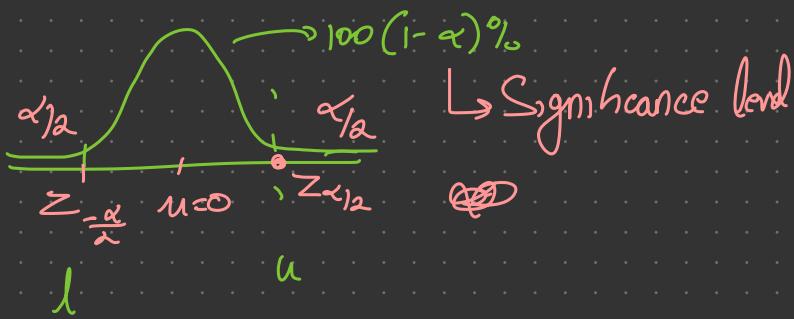
$$Z = \frac{\bar{x} - E(\bar{x})}{\sigma_{\bar{x}}}$$



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

$$P\left[l \leq z \leq u\right] = 1 - \alpha$$

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



How to get l, u ?

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

$$-z_{\alpha/2} \leq \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}$$

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

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

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

↑

Confidence Interval

$100(1-\alpha)\%$

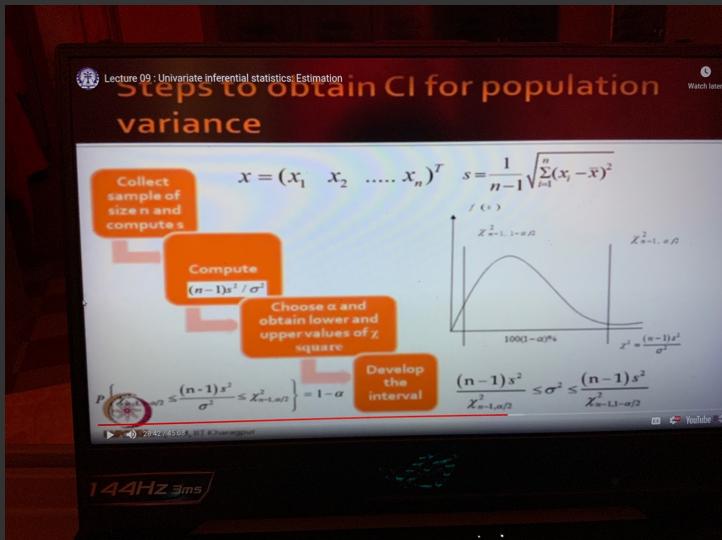
NOTE:

When we say CI, it's about population parameters & not sample statistics

If we are using t-distribution,

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

$$\bar{x} - t_{n-1} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{n-1} \frac{s}{\sqrt{n}}$$

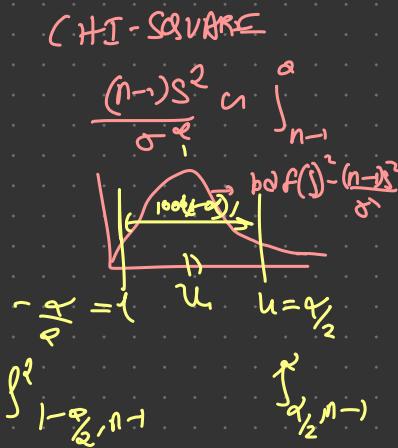


Confidence Interval for

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

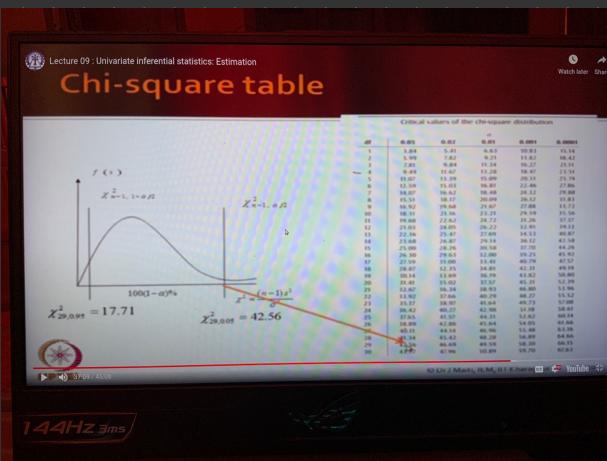
$$\frac{l^2}{1-\frac{\alpha}{2}, n-1} \leq \frac{(n-1)s^2}{\sigma^2} \leq \frac{u^2}{\frac{\alpha}{2}, n-1}$$

$$\frac{(n-1)s^2}{l^2_{1-\frac{\alpha}{2}, n-1}} \leq \sigma^2 \leq \frac{(n-1)s^2}{u^2_{\frac{\alpha}{2}, n-1}}$$





Chi-square table



Lecture

CI for $\mu_1 - \mu_2$

$$(\bar{x}_1 - \bar{x}_2) - Z_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \leq \mu_1 - \mu_2 \leq (\bar{x}_1 - \bar{x}_2) + Z_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

☞ s_1, s_2 must be known
 n_1, n_2 for normal population.

If, $\begin{cases} s_1 = s_2 = \sigma \\ \downarrow \end{cases}$ Population variances are same

Another estimation

First find $\hat{\sigma}_{\text{pooled}}$

$$\begin{cases} \hat{\sigma}_{11} = \hat{\sigma} \\ \hat{\sigma}_{22} = \hat{\sigma} \end{cases}$$

$$\hat{\sigma}_p^2 = \sigma_{\text{pooled}}^2 = \frac{(n_1 - 1)\hat{\sigma}_1^2 + (n_2 - 1)\hat{\sigma}_2^2}{n_1 + n_2 - 2} = \hat{\sigma}^2$$



$$\sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}} = \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = \hat{\sigma}_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

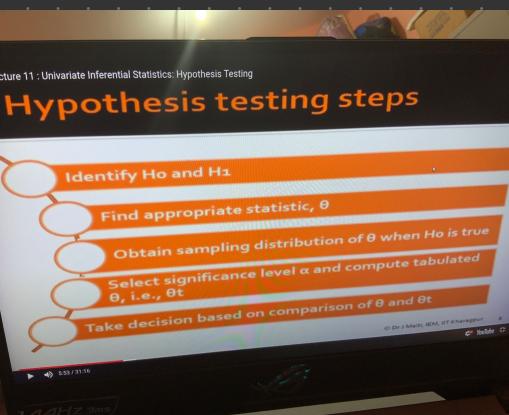
$$(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2} \left[\hat{\sigma}_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right] \leq \mu_1 - \mu_2 \leq (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2} \hat{\sigma}_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Lecture 11

Hypothesis testing:

Statement that is yet to be proven

Population parameters



Introduction

- ① → ▪ A hypothesis is a statement that is yet to be proven
- ② → ▪ H₀: Null hypothesis
 - An assertion about the value of a population parameter
 - Hold as true unless statistical evidence concludes otherwise
- H₁: Alternative hypothesis
 - Negation of H₀.

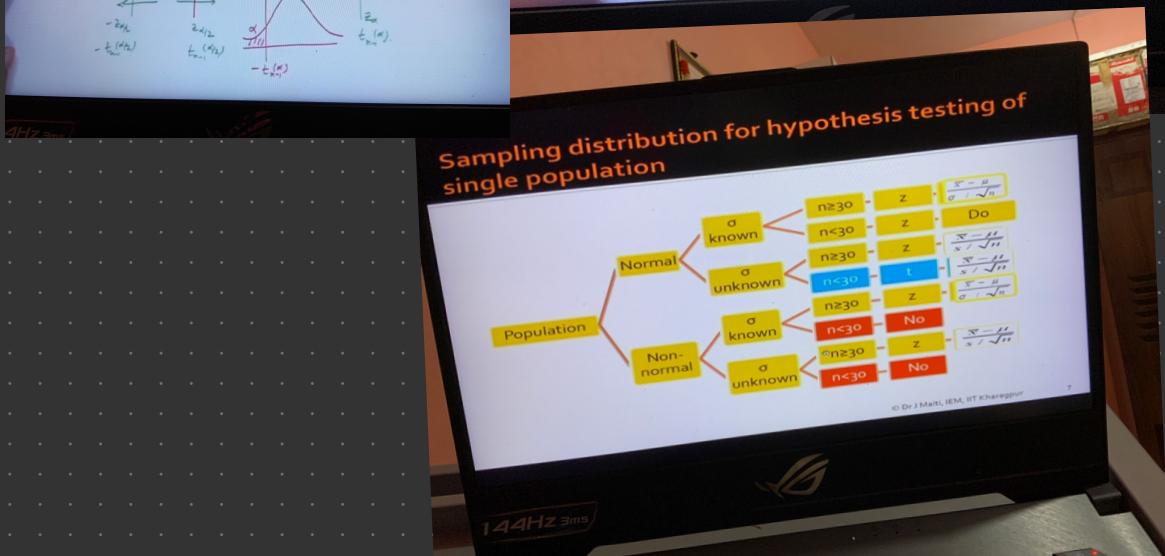
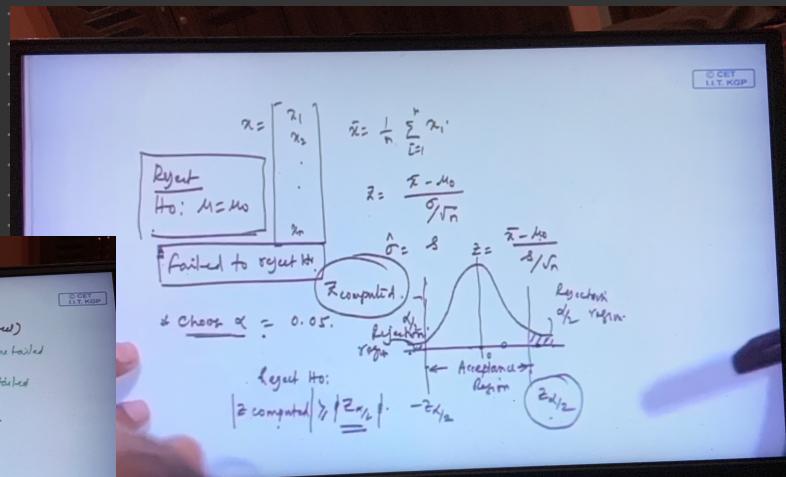
Dr J. Math, IITM, IIT Kharagpur

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

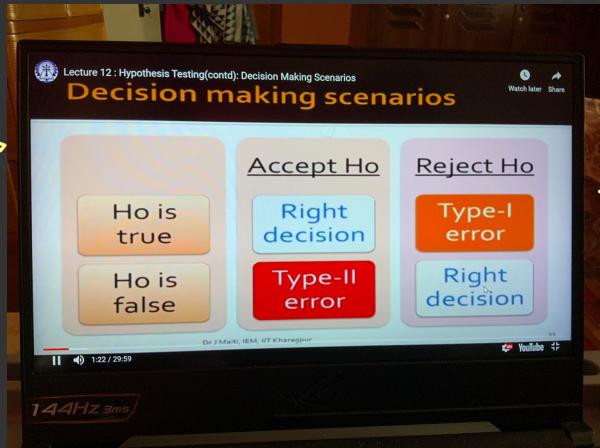
$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

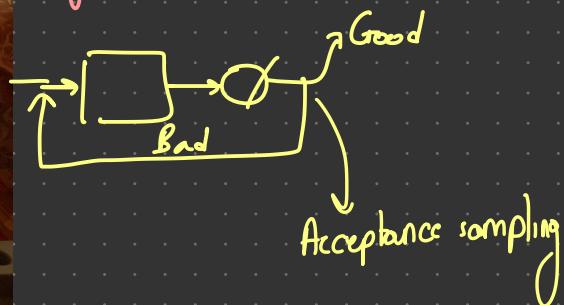
Assuming H_0 is true
Statistic for which
we need to know
the mean of



Decision Making Scenarios



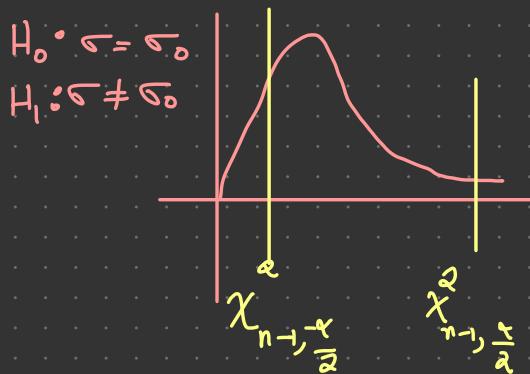
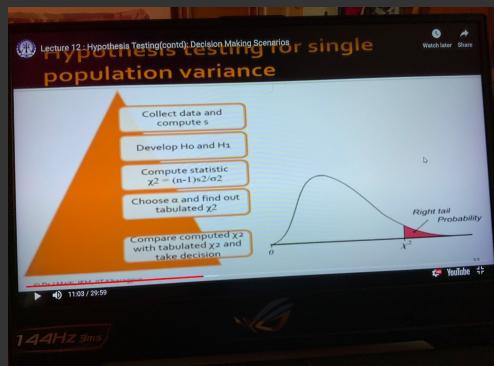
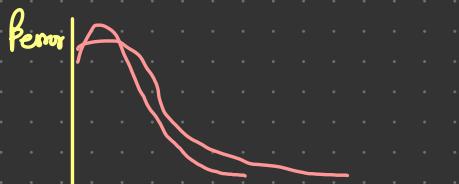
Physical Interpretation



Operating On characteristic Curve [OC Curve]

$$\beta \quad 1 - \beta: \text{Power of test}$$

Not able to identify mean diff is too small



Hypothesis testing for equality of two population means

Collect samples of sizes n_1 and n_2 from populations 1 and 2 , respectively

Compute mean difference and its variance

$$\mu_{\bar{x}_1 - \bar{x}_2} = E(\bar{x}_1 - \bar{x}_2) = E(\bar{x}_1) - E(\bar{x}_2) = \mu_1 - \mu_2$$

$$\sigma_{\bar{x}_1 - \bar{x}_2}^2 = V(\bar{x}_1 - \bar{x}_2) = V(\bar{x}_1) + V(\bar{x}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Compute statistic

Find out appropriate sampling distribution

Test hypothesis

$$\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}} \sim N(0, 1)$$

- For normal populations with known σ_1 and σ_2 but large sample size
- For non-normal populations with known σ_1 and σ_2 but large sample size

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

Hypothesis testing for the equality of two population variances

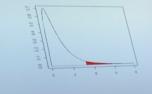
Collect samples of sizes n_1 and n_2 from populations 1 and 2 , respectively

$$\frac{s_1^2}{s_2^2} / \frac{\sigma_1^2}{\sigma_2^2} = \frac{(n_1-1)s_1^2}{(n_2-1)s_2^2} / \frac{\sigma_1^2}{\sigma_2^2} \sim F_{n_1-1, n_2-1}$$

Compute sample variances and appropriate F statistic

Find out appropriate sampling distribution

Test hypothesis



Books of Fouad & Miller

References

- Aczel A D (2010). Complete business statistics. Tata McGraw Hill, Sixth Edition, 820p.

Hypothesis testing for the equality of two-population means: Special case

Collect samples of sizes n_1 and n_2 from populations 1 and 2 , respectively

$$\sigma_1^2 = \sigma_2^2 = \sigma^2$$

Compute mean difference and its variance

$$\mu_{\bar{x}_1 - \bar{x}_2} = E(\bar{x}_1 - \bar{x}_2) = E(\bar{x}_1) - E(\bar{x}_2) = \mu_1 - \mu_2$$

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}$$

Compute statistic

$$t_{n_1+n_2-2} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Find out appropriate sampling distribution

Test hypothesis