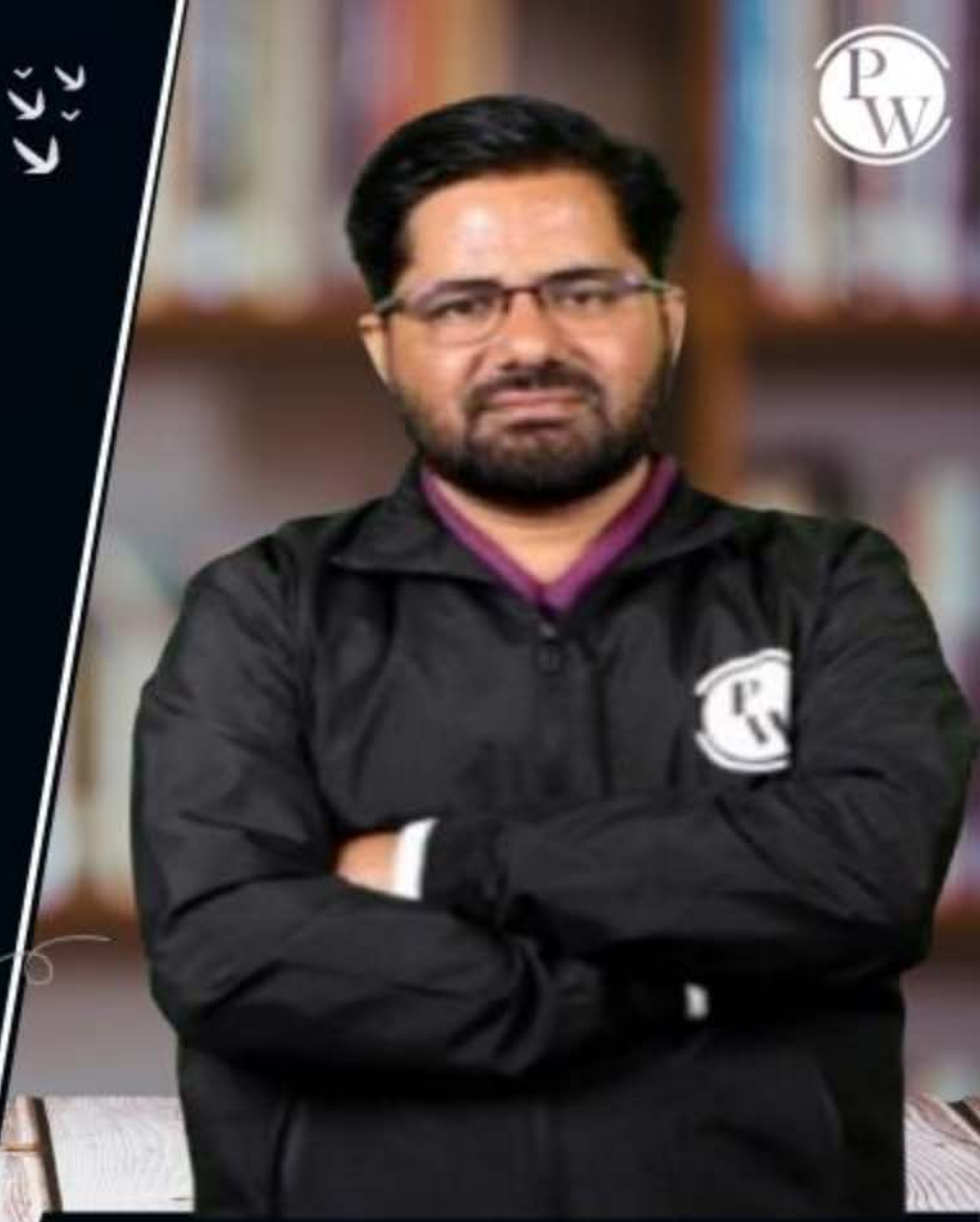
Computer Science & DA

Probability and Statistics

Sampling Theory & Distribution

Lecture No. 03



By- Dr. Puneet Sharma Sir

Recap of previous lecture

Topic Z - test

$SE(\bar{x}) = \dfrac{\sigma}{\sqrt{n}}$,

$SE(\tilde{p}) = \sqrt{\dfrac{\tilde{p}\tilde{q}}{n}}$

$\tilde{p} = \dfrac{x}{n} = \dfrac{\text{success}}{\text{Total}}$

$\tilde{q} = 1 - \tilde{p}$

$\bar{x} - 3SE(\bar{x}) \le \mu_0 \le \bar{x} + 3SE(\bar{x})$

$\tilde{p} - 3SE(\tilde{p}) \le p_0 \le \tilde{p} + 3SE(\tilde{p})$

Topics to be Covered

Topic t - Distribution

Topic :t - Distribution

RECAP of Z-test :-

Type 1 : Significance of population Mean :

$H_0 : \boxed{\mu = \mu_0}$, $H_1 : \mu \neq \mu_0$

$$\boxed{Z = \frac{\bar{x} - \mu_0}{\sqrt{\sigma^2 / n}}}$$

Type 2 :- Significance of Diff b/w two Pop Mean :-

$H_0 : \boxed{\mu_1 = \mu_2}$, $Z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

$H_1 : \mu_1 \neq \mu_2$

Type 3 :- Significance of population prop :-

$H_0 : \boxed{p = p_0}$, $H_1 : p \neq p_0$

$$Z = \frac{x - \mu}{\sigma} \approx \frac{x - np_0}{\sqrt{np_0q_0}} = \frac{\left(\frac{x}{n}\right) - p_0}{\sqrt{\frac{p_0q_0}{n}}}$$

$$\boxed{Z = \frac{\tilde{p} - p_0}{\sqrt{\frac{p_0q_0}{n}}}}$$

Type (4) Significance of Diff b/w two Pop Prop :-

$H_0 : p_1 = p_2$

$H_1 : p_1 \neq p_2$

$$\boxed{Z = \frac{\tilde{p}_1 - \tilde{p}_2}{\sqrt{\frac{\tilde{p}\tilde{q}}{n_1} + \frac{\tilde{p}\tilde{q}}{n_2}}}}$$

$\tilde{p} = \frac{x_1 + x_2}{n_1 + n_2}$

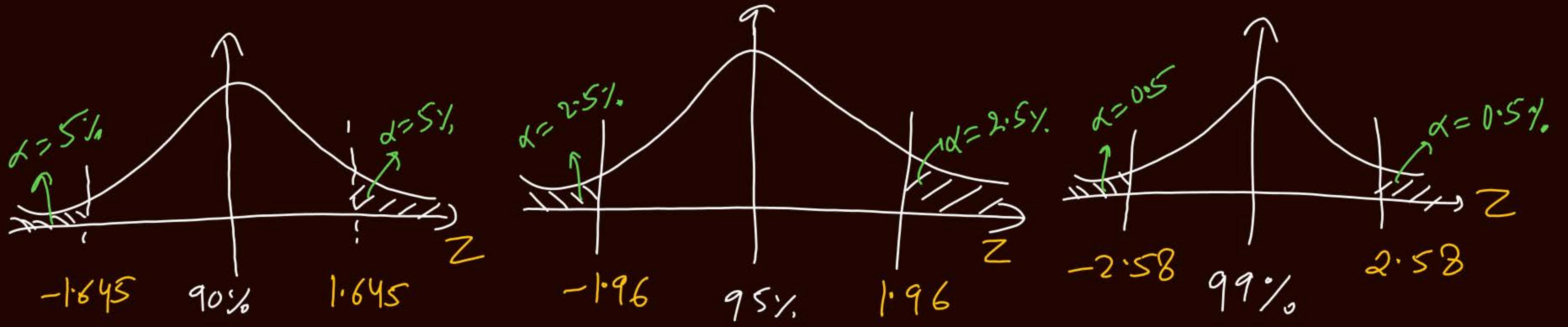
$\tilde{p}_1 = \frac{x_1}{n_1}$, $\tilde{p}_2 = \frac{x_2}{n_2}$

Type (4) Significance of Difference b/n two population proportion —

$\tilde{p}_1 = \frac{x_1}{n_1}$; $\tilde{p}_2 = \frac{x_2}{n_2}$, Now common proportion is, $\tilde{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{n_1\tilde{p}_1 + n_2\tilde{p}_2}{n_1 + n_2}$

$H_0 : \boxed{p_1 = p_2}$; $H_1 : p_1 \neq p_2$ & $\tilde{q} = 1 - \tilde{p}$.

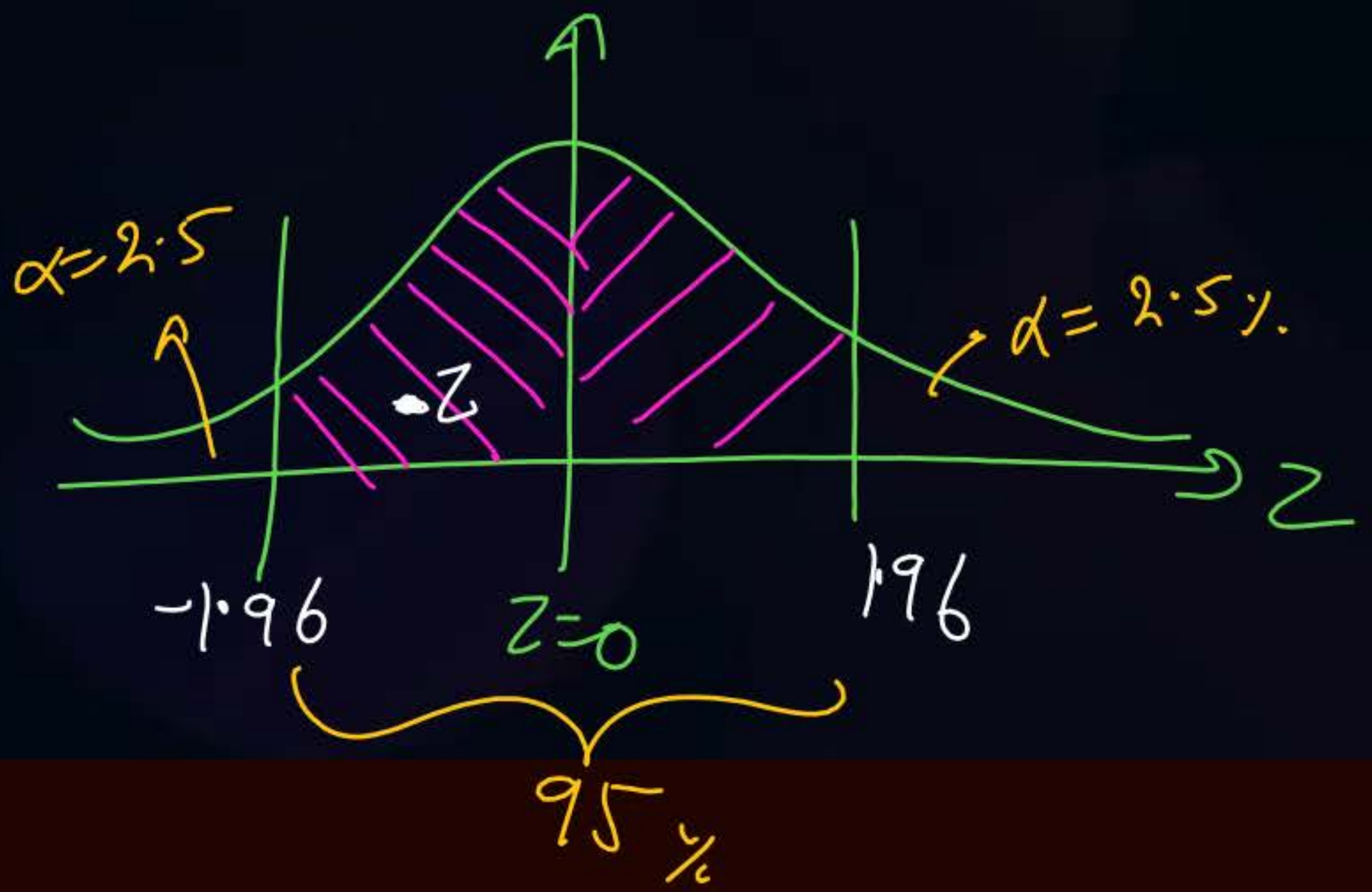
$$Z = \frac{\tilde{p}_1 - \tilde{p}_2}{\sqrt{\frac{\tilde{p}\tilde{q}}{n_1} + \frac{\tilde{p}\tilde{q}}{n_2}}}$$


α = 5%
α = 5%
Z
-1.645
90%
1.645
α = 2.5%
α = 2.5%
Z
-1.96
95%
1.96
α = 0.5
α = 0.5%
Z
-2.58
99%
2.58

#Q. A sample of 400 members has a mean = 4 where sample in taken from normal population with unknown mean and standard deviation 2.6 can we say that population mean is 4.2 with 5% level of significance. It is given that for two tailed test, $Z_\alpha = 1.96$ for $\boxed{\alpha = 0.025}$.

Type 1

$n = 400$, $\bar{x} = 4$, $\sigma = 2.6$, $H_0 : \mu_0 = 4.2$, $H_1 : \mu_0 \neq 4.2$ (Two Tailed)



$$Z = \frac{\bar{x} - \mu_0}{\sqrt{\sigma^2/n}} = \frac{4 - 4.2}{\sqrt{(2.6)^2/400}} = -1.538$$

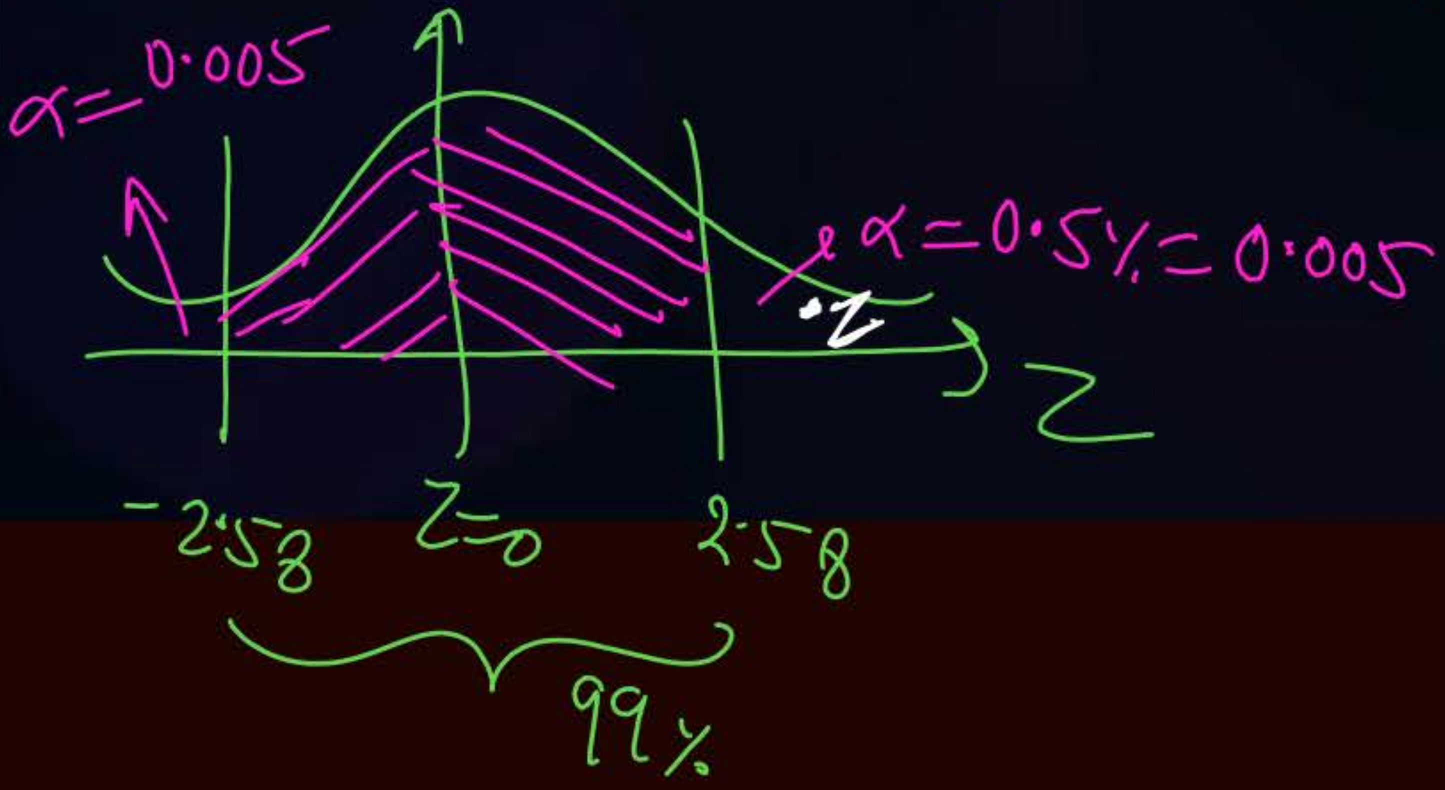
$\because |Z| < Z_\alpha(5\%)$ Hence H_0 accepted.

#Q. While calculating the average monthly income of a family in a town a sample of 81 families was taken. The mean income and ~~standard deviations~~ of these 81 families were found to be 4108 Rs and 24 Rs respectively shown that the assumption "average income of family in a town in 4100 Rs" is not reasonable for 1% level of significant if $Z_\alpha = 2.58$.

$= \sigma$

(ii) Also find the most probable limits for average income.

$\bar{x} = 4108, \ \sigma = 24, \ n = 81, \ H_0 : \mu_0 = 4100, \ H_1 \neq 4100$



$$Z = \frac{\bar{x} - \mu_0}{\sqrt{\sigma^2 / n}} = 3$$

$\because |Z| > Z_\alpha (1\%)$

i.e. H_0 is Rejected

i.e. $\mu \neq 4100$ Rs

#Q. The mean of two samples of 1000 and 2000 members are 67.5 and 68.0 inches respectively can the sample be regarded as drawn from the same population of standard deviation 2.5 inches take [illegible]

Type2 i.e. Z_α (0.05) = 1.96

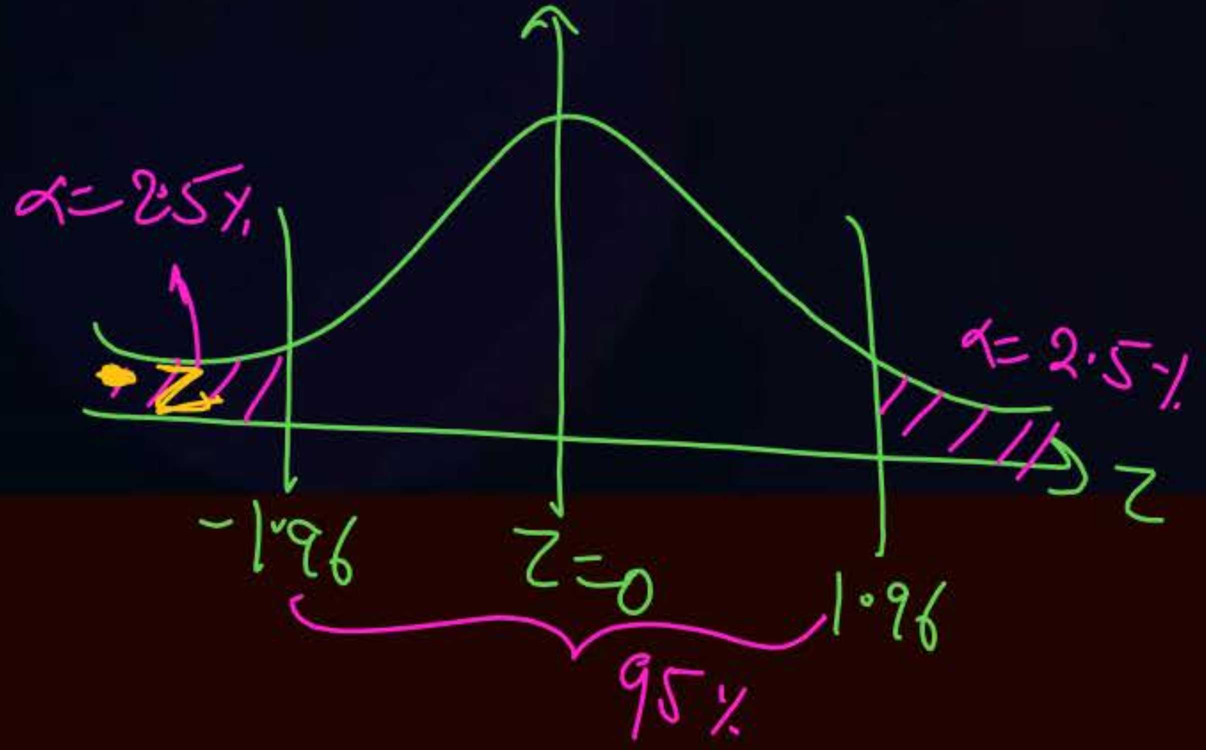
$n_1 = 1000$, $n_2 = 2000$

$\bar{x} = 67.5$, $\bar{y} = 68$

$\sigma_1 = \sigma_2 = 2.5$

$H_0: \boxed{\mu_1 = \mu_2}$, $H_1: \mu_1 \neq \mu_2$

$$Z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = -5.16$$



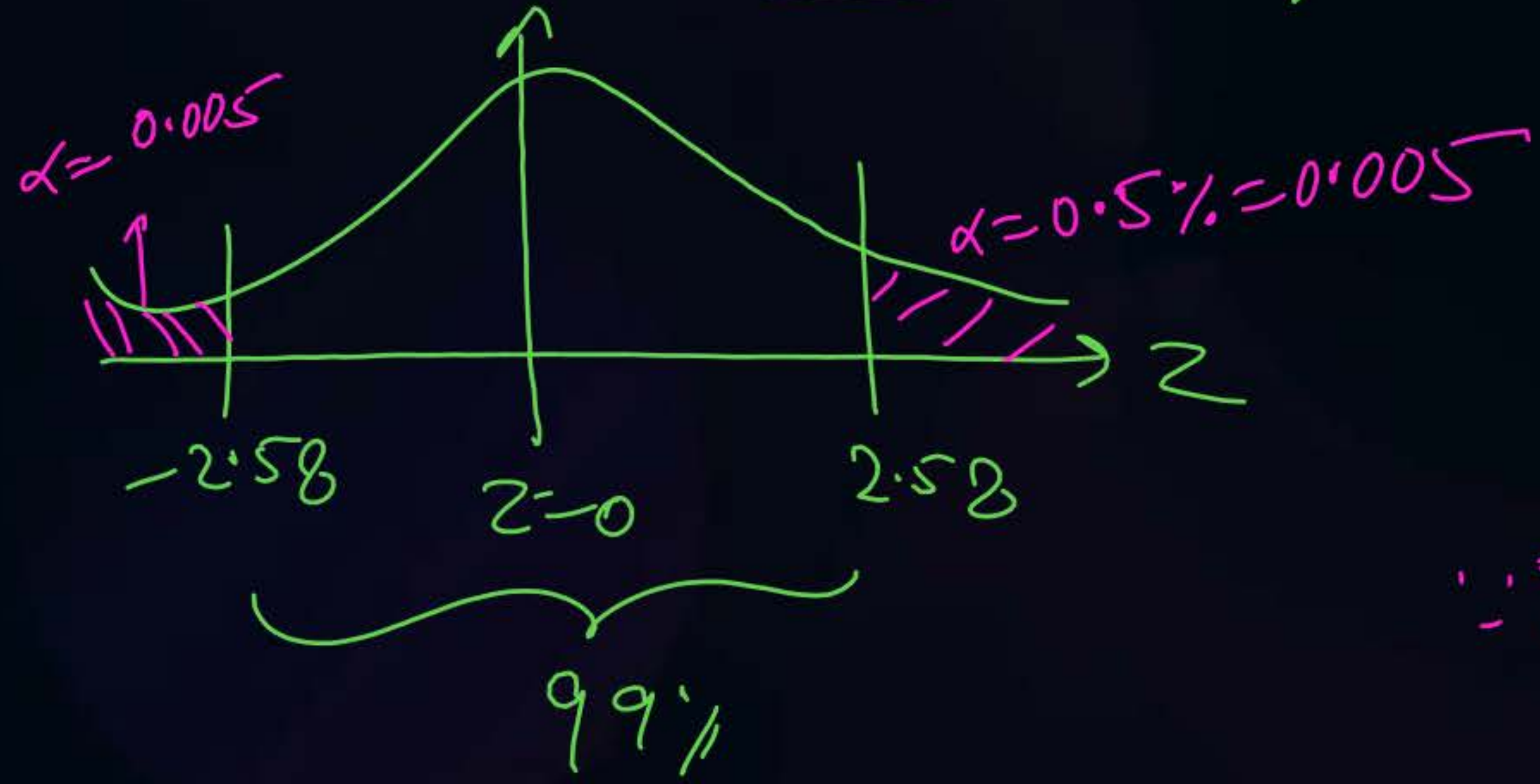
∵ Z lies in R-Region so H_0 is Rejected & H_1 is Accepted.

#Q. If means of two samples (of same size 400) are 250 and 220 with their standard deviation 40 and 55 respectively then test $\boxed{H_0: \mu_1 = \mu_2}$ against $H_1: \mu_1 \neq \mu_2$ at 1% level of significance

i.e. $Z_\alpha (0.01) = 2.58 \Rightarrow \alpha = 0.01 = 1\%$

$n_1 = n_2 = 400, \quad \bar{x} = 250, \quad \bar{y} = 220,$

$\sigma_1 = 40, \quad \sigma_2 = 55$



$$Z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = 8.82$$

$\because$ Z lies in R. Region so H_0 : Rejected

H_1 : Accepted

#Q. A coin was tossed 400 times and head tossed up 210 times discuss whether coin is unbiassed or not. $Z_\alpha(10\%) = 1.645$

Sol: $x = \{\text{Number of Head}\} \rightarrow$ success.

Pop. Prop (if Dice is unbiassed) $p_0 = \frac{1}{2} = 0.5$

$H_0 : \boxed{p_0 = 0.5}$ (unbiassed), $H_1 : p_0 \neq 0.5$ (Biased)

$q_0 = 0.5$

Sample Prop of Head, $\tilde{p} = \frac{x}{n} = \frac{210}{400} = 0.525$

$n = 400$

$$Z = \frac{\tilde{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.525 - 0.500}{\sqrt{\frac{0.5 \times 0.5}{400}}} = 1$$

Hence H_0 is accepted i.e. $p_0 = 0.5$ or coin is unbiassed.

(ii) Also find the Most probable limits for p_0 ?

$$\tilde{p} - 3\,SE(\tilde{p}) \le p_0 \le \tilde{p} + 3\,SE(\tilde{p})$$

$$? \le p_0 \le ?$$

$$SE(\tilde{p}) = \sqrt{\frac{\tilde{p}\tilde{q}}{n}} = \sqrt{\frac{0.525 \times 0.475}{400}}$$

#Q. A Die was thrown 9000 times and 1 or 6 was obtained 3240 times can we say that die was unbiased for 1% level of significance?

$Z_\alpha\ (1\%) = 2.58$

$n = 9000$, $x = \{$Number of times $\underbrace{\text{1 or 6 is occurring}}_{\text{success}}\}$

$\bar{p} = \frac{x}{n} = \frac{3240}{9000} = 0.36$

$p_0 =$ Prob of getting 1 or 6 $= \frac{2}{6} = \frac{1}{3} = 0.33$

$H_0 : \boxed{p_0 = 0.33}$ (unbiased), $H_1 : p_0 \neq 0.33$

& $q_0 = 0.67$

$$Z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = 6.04$$

∵ Z lies in R-Region so H_0 is Rejected & H_1 = Accepted is BIASED

#Q. In a town, 350 out of 600 person were found to be vegetarian. On the basis of this data, can we say that majority of population in the town in vegetarian at 5% level of significant

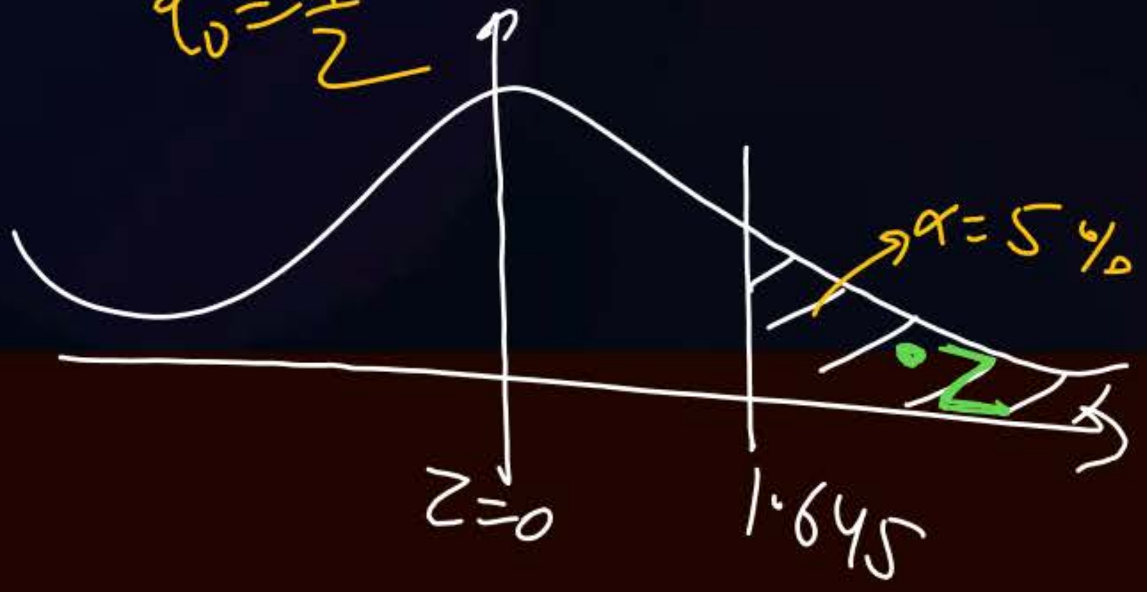
Given that for right failed $Z_{\alpha}(0.05) = 1.645$

$x = \{\underbrace{\text{No. of Vegetarian Person}}_{\text{success}}\}$

$n = 600$, $x = 350$, $\bar{p} = \frac{x}{n} = \frac{350}{600} = 0.58$

$H_0 : \boxed{p_0 = \frac{1}{2}} = 0.5$, $H_1 : p_0 > \frac{1}{2}$ (Right Tailed test)

$q_0 = \frac{1}{2}$



$$Z = \frac{\bar{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.58 - 0.50}{\sqrt{\frac{0.50 \times 0.50}{600}}}$$

$$= 3.91$$

Z lies in R-Region so H_0 is Rejected & H_1 is Accepted

i.e. Majority of Population is Vegetarian.

#Q. There are two sample of 1000 person from city A and 800 person from city B, 400 person from each sample are found to be ~~what~~ consumers of wheat. Can we say that number of wheat consumers from city A and city B differs significantly in desired % of confidence (i.e. 99.7%)

Sol: $n_1 = 1000$, $n_2 = 800$, $x = \{$ No. of wheat consumers $\} \rightarrow$ success

$x_1 = 400$, $x_2 = 400$,

$\tilde{p}_1 = \frac{x_1}{n_1} = \frac{400}{1000} = 0.4$,

$\tilde{p}_2 = \frac{x_2}{n_2} = \frac{400}{800} = 0.5$,

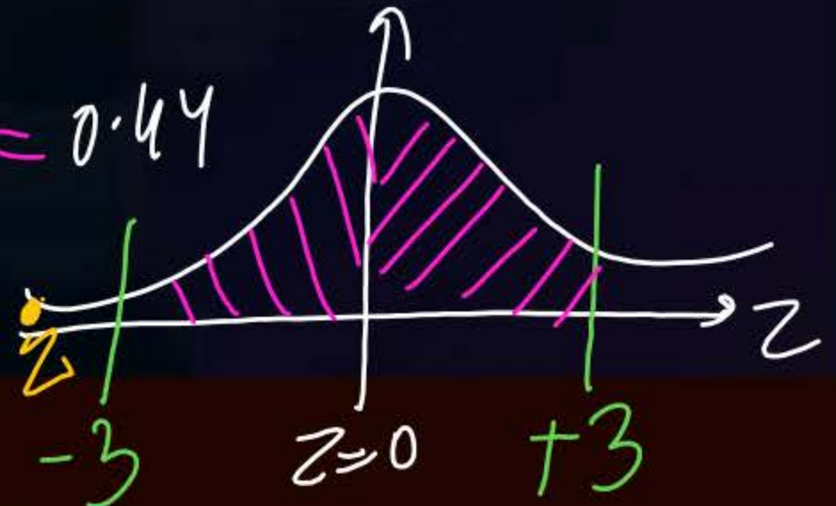
$$Z = \frac{\tilde{p}_1 - \tilde{p}_2}{\sqrt{\frac{\bar{p}\bar{q}}{n_1} + \frac{\bar{p}\bar{q}}{n_2}}} = \frac{0.4 - 0.5}{\sqrt{\frac{0.44 \times 0.56}{1000} + \frac{0.44 \times 0.56}{800}}} = -4.25$$

where $\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{400 + 400}{1000 + 800} = 0.44$

& $\hat{q} = 0.56$

H_0 : $\boxed{p_1 = p_2}$ i.e No Diff b/w wheat consumers in A & B

H_1 : $p_1 \neq p_2$ i.e there is a diff " " " " "



Hence H_0 is Rejected & H_1 is accepted.

#Q. A machine produced 20 defective articles in a lot of 400 and after overhauling it produced 20 defective articles in a lot of 300. Has the machine improved at $\alpha = 0.01$.

HW

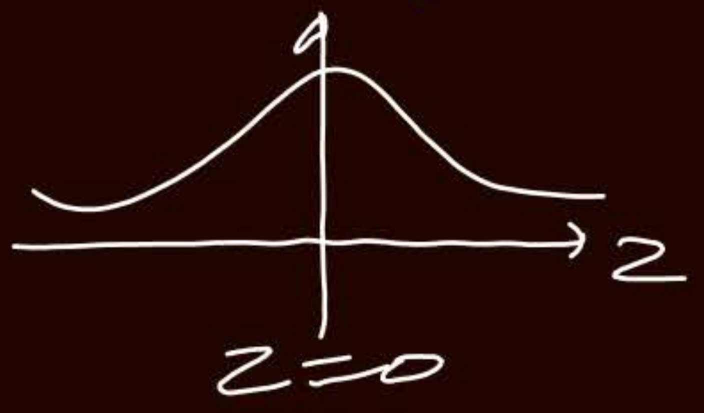
Type 4

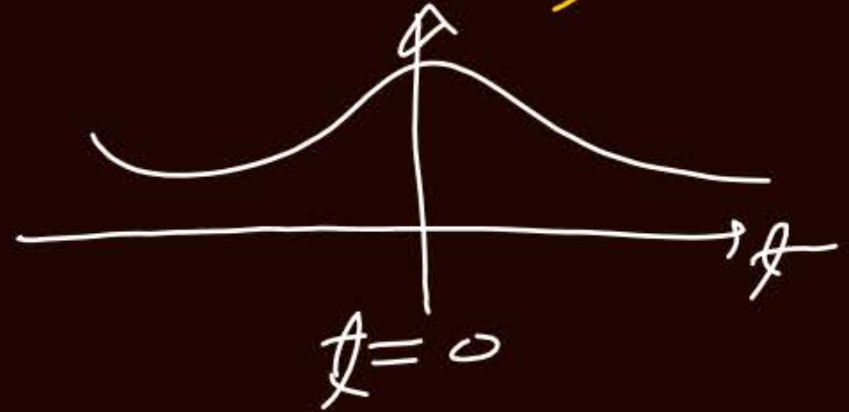
For right failed $Z_{\alpha}(0.01) = 2.33$

(Machine has not been improved)

Student's t-test (Small sample test) (i.e. for $n < 30$)

* Z test :- Large sample test (i.e. for $n > 30$)





Pop Variance $(\sigma^2) = \frac{\sum (x-\bar{x})^2}{N}$

Sample Variance $(s^2) = \frac{\sum (x-\bar{x})^2}{(n-1)}$

Degree of freedom :- $\boxed{\text{Df = Number of observation − No. of conditions imposed on them}}$

i.e. Max number of Ind values that have freedom to vary is called D-f.

eg: If we are selecting 5 Nos whose sum is 50
i.e. $Df = 5-1 = 4$
∵ 5^{th} No = 50 − Sum of 4 Nos

eg: If we are selecting 40 persons whose average wt is 65 kg and SD is 9 kg then
$Df = 40 - 2 = 38$

- for Large Degree of freedom, t-test converts into Normal Distribution
- " " " " , Chi-Sq test " " Normal Distribution
- for two sample test ; $Df = (n_1 + n_2) - 2$
- with the Help of t-test we can solve following types of Questions →

Type I :→ Significance of Population Mean →

$H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$

$$\boxed{t = \frac{\bar{x} - \mu_0}{\sqrt{S^2/n}}}$$ where $S^2 = \frac{\sum (x - \bar{x})^2}{n-1}$

Type II → Significance of Difference b/n two population Mean :→

#Q. A random sample of 16 values from normal population has mean of 41.5 cm and sum of squares of deviation from mean is 135 cm^2. ~~can we say that the popular mean is 43.5 cm?~~ can we say that the population mean is 43.5 cm? with 5% level of significance.

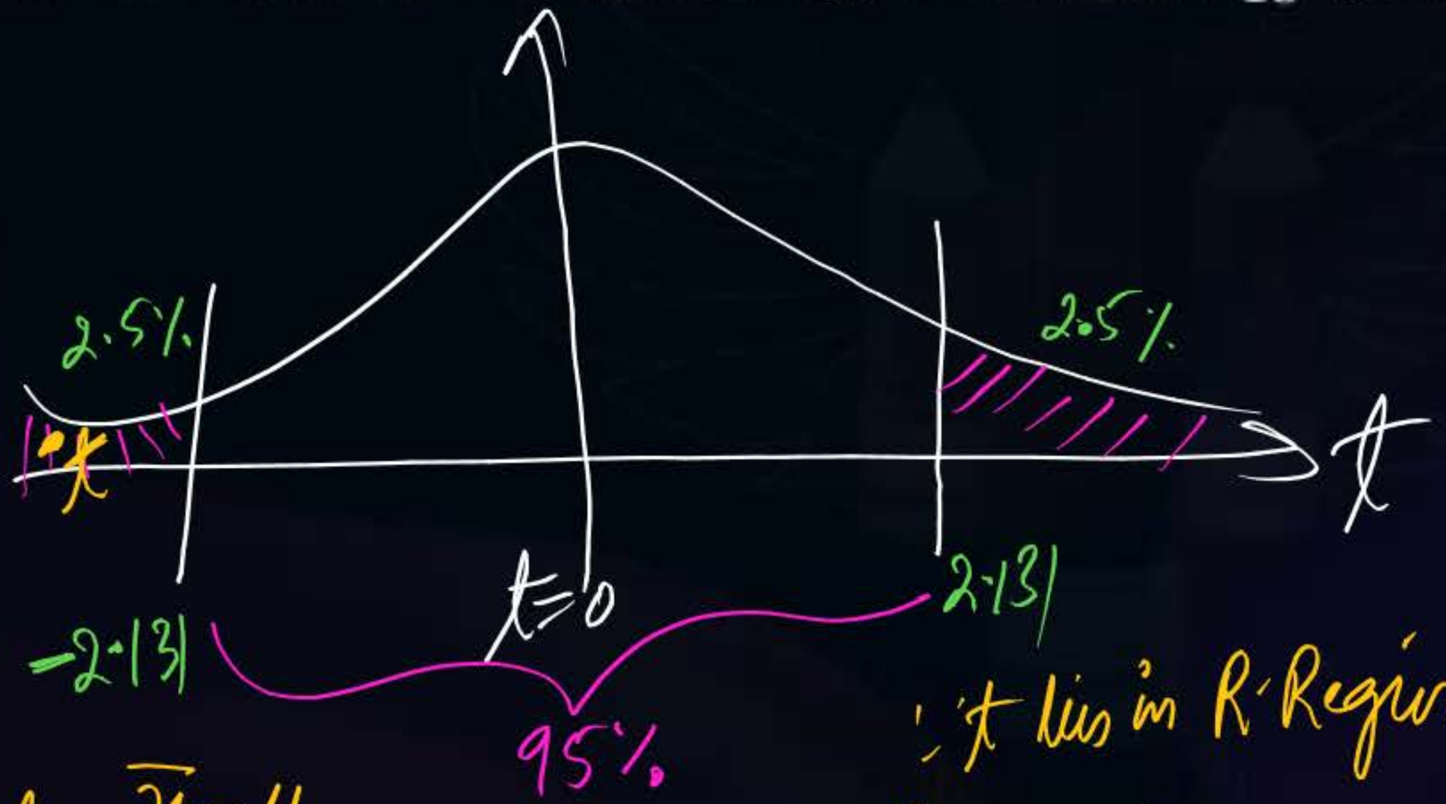
(ii) Also find the 95% and 99% confidence for μ it is given that $t_{15}(0.05) = 2.131$ and $t_{15}(0.01) = 2.94$

$n = 16$, $\bar{x} = 41.5$, $\sum(x-\bar{x})^2 = 135$

$s^2 = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{135}{16-1} = 9$

$H_0 : \boxed{\mu_0 = 43.5}$, $H_1 : \mu_0 \neq 43.5$

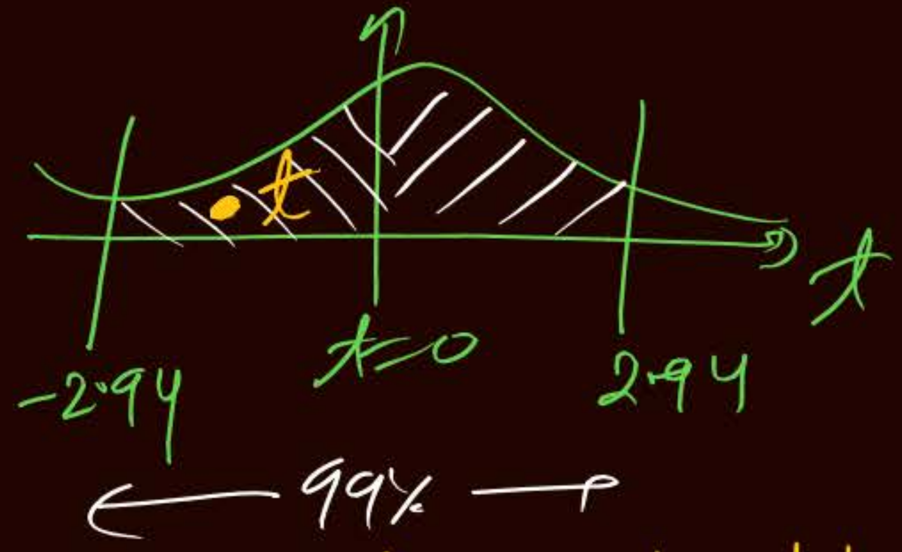
& $Df = n-1 = 16-1 = 15$



$t = \frac{\bar{x}-\mu_0}{\sqrt{s^2/n}} = -2.67$

∵ t lies in R. Region so Neglect H_0 & Accept H_1 ie $\mu_0 \neq 43.5$

Note: Also check Validity of H_0 or H_1 at 99% Confidence limit?



$$t = \frac{\bar{x} - \mu_0}{\sqrt{s^2/n}} = -2.67$$

$\because$ t lies in Acceptance Region so H_0 accepted

i.e. $\boxed{\mu_0 = 43.5}$

Most probable limits for 95%

$$\bar{x} - t_{15}(0.05)SE(\bar{x}) \le \mu \le \bar{x} + t_{15}(0.05)SE(\bar{x})$$

$$41.5 - 2.131(0.75) \le \mu \le 41.5 + 2.131(0.75)$$

$$39.99 \le \mu \le 43.09$$

Ans

(ii) $SE(\bar{x}) = \frac{s}{\sqrt{n}} = \frac{3}{\sqrt{16}} = \frac{3}{4} = 0.75$

(iii) $\bar{x} - t_{15}(1\%)SE(\bar{x}) \le \mu \le \bar{x} + t_{15}(1\%)SE(\bar{x})$

$$39.2 \le \mu \le 43.7$$

#Q. Average height of 10 student is a school is observed as 67 inches with sum of the squares of deviations from central value is 88. can we say that average height of student in a school is 65 inches. It is given that, value of t at 5% level of significance with 9 degree of freedom is 2.262.

HW

#Q. A machine produces washers of thickness 10mm. A sample of 10 washers has an average thickness of 9.52 mm with SD of 0.6 mm. Find out 't'.

S.D

$n = 10$, $\bar{x} = 9.52$, $s = 0.6$

$\mu_0 = 10$

$$t = \frac{\bar{x} - \mu_0}{\sqrt{s^2/n}} = \frac{9.52 - 10}{\sqrt{(0.6)^2/10}} = -2.529$$

THANK - YOU