Till now

→ Pdf

→ Hypothesis testing

Write the mathematical eqn, H_0 → whichever has equality.

whatever is widely accepted.

Sachin → the avg age of Deloitte employee is 45 years.
(claim/statement/hypothesis)

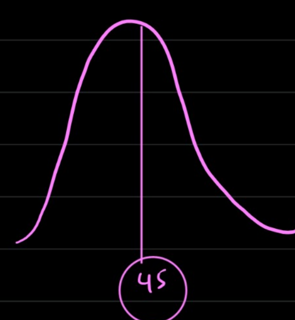
H_0 : $\mu_{age} = 45$ years

H_A : $\mu_{age} \neq 45$ years

greater than 45 years

$\mu_{age} > 45$ → H_A.

$\mu_{age} \leq 45$ → H_0



Hypothesis

Mathematical

→ equality symbol should be there in H_0

Practical scenarios (whatever is accepted in Society)

① Social trial → Rahul ← Agrim

Court (Judge)
(Vatsalya)

Everyone is innocent in the court until proven guilty.

H_0 : Rahul is innocent

② ABC medical test for Cancer.

H_0 : ABC doesn't have cancer.
H_A : " has cancer.

* Level of Significance & confidence.

* We can not be 100% sure (confident) to conclude something for the population using sample.

* If a pharma company wants to know if a vaccine will work or not.

↓

1000 sample

↓

950 → people is well / cured with this vaccine

✓ Vaccine works 95% of the time.

↓

95% ⟹ Some threshold

> Out of 100 times, if I conduct a test how many times same type of conclusion can be made.

Scen-1

* Vaccine works 95% of time.

⇓

Group-1 → 950 people (atleast)

⇓

5%

Scen-2

	G_1	G_2	G_3	G_4	G_5	---	G_{100}
1000	1000	1000	1000	1000	100		

* Among these 100 groups 95 groups atleast should be there where more 950 people are cured with vaccine.

⇓

5%

Vaccine works 95% of the time.

↓

95% ⟹ Some threshold

⟹ Out of 100 times, if I conduct a test how many times same type of conclusion can be made.

→ How much margin of error you are ready to accept.

* If I say I am ready to accept 5% margin of error then it simply out of 100 experiment 95% of the time. vaccine will cure and rest of 5% time it will not cure.

→ 5% → level of Significance.

⇓

The percentage of risk involved in testing. alpha.

$\alpha \rightarrow$ risk. (5%)

$1-\alpha \rightarrow$ Confidence.

$100 - 5\% = 95\%$

$\alpha = 1, 5, 10\%$

→ C.I = 99%, 95%, 90%

① Framing of hypothesis

② alpha (margin or error)

③ What type of test that you will use.

→ Samples statistics | test statistics

↓

→ Z statistics

→ t statistics

→ F - statistics

→ Ch-sq statistic

① S.S ≥ 30

→ σ pop → SND

↓

Z distn

⇓

Z test

② S.S < 30 or σ pop not given

→ t distn

→ t test.

③ Variance of two sample/or more than two.

→ F distn

④ test Critical

- ↳ Z critical
- ↳ t critical
- ↳ F critical
- ↳ Chi square critical.

↳ F test, ANOVA

④ Two categorical

↳ Chi square dist^n

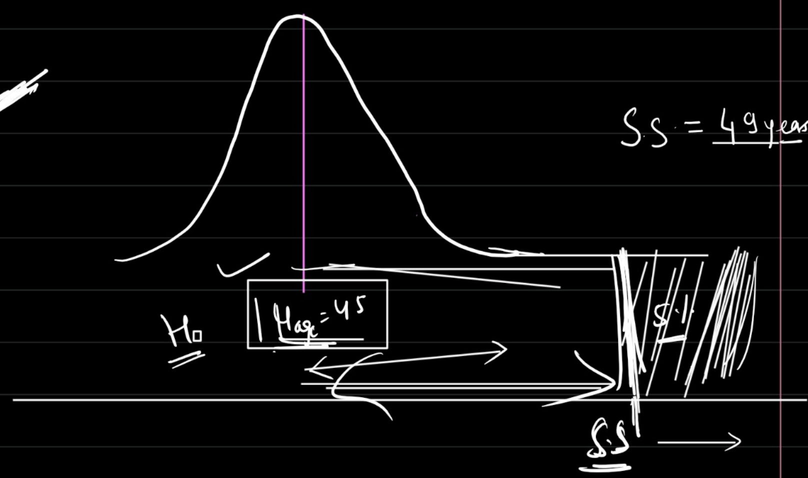
↳ Chi-square test.

Sachin's $H_0: \mu_{age} = 45$ years

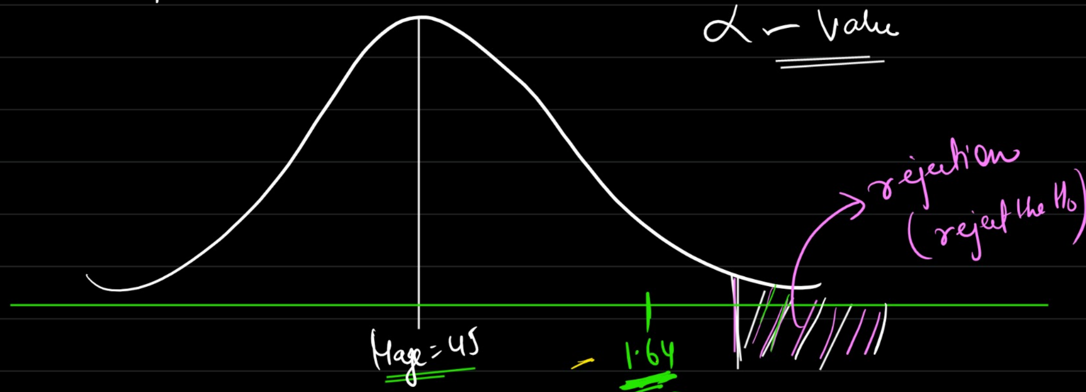
$H_A: \mu_{age} \neq 45$ years

S.S = 49 years

* How far your S.S should be there, so that you can reject H_0



α – value



$\alpha = 5\%$

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

↓

How much std dev it is away from mean value.

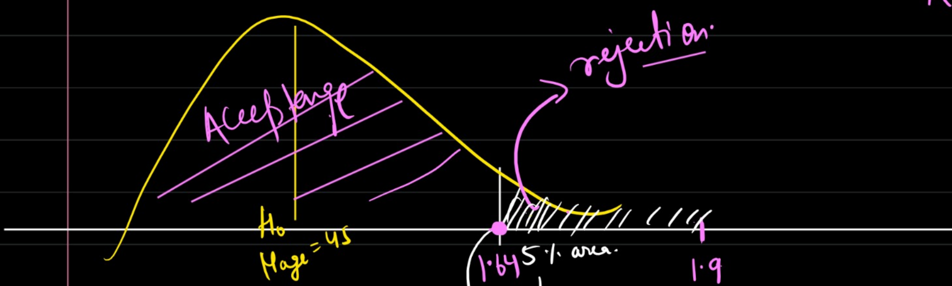
two approach

- Convert α → 5% into Zscore. → Z critical
  - Z table → Zscore & Corresponding Prob value will be give
- Convert Sample statistics to Prob value

Step-5: Conclusion

$\alpha = 5\%$

Scen-1



$S.S = \underline{1.9}$
(Zscore)

5% area → convert to to scale Zscore.

⇓

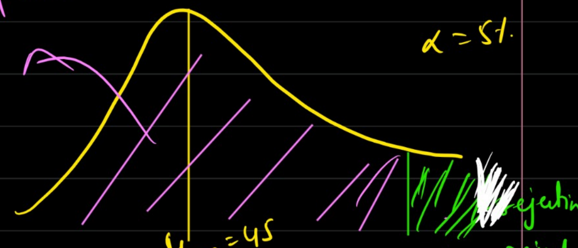
Statistable = 5% area = 1.64

$S.S > 1.64 \Rightarrow$ rejection region
⇒ you reject the H_0

if $S.S < 1.64$ → fail to reject the H_0

Scen 2

$\alpha = 5\%$



$\mu_{age} = 45$

$S.S = 1.9$

↓

Prob value corresponding to sample score.

↓

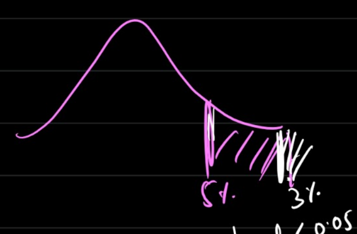
0.03

⇓

p value
p value ≤ 0.05
↓
You reject the H_0

5% → reject
$\frac{5}{100} = 0.05$

0.05



↓

Pvalue > 0.05 → fail to reject H_0

* Pr value – The p-value is the probability value, calculated from a test statistics.
→ Use to decide whether to reject a H_0 or not.

* One tail test, two tail test.

α → whole experiment

$\alpha = 0.05$

$H_0 : \mu_{age} = 45$
$H_A : \boxed{\mu_{age} \neq 45}$

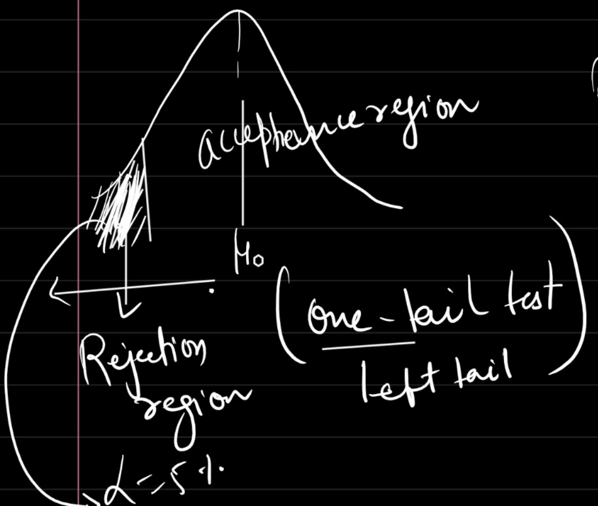
$\neq$ → Alternate Hypothesis
↳ two tail test.
→ alpha/2

two tail test

Scen-1

$H_0 : \mu_x = \mu_0$

$H_A : \mu_x < \mu_0$

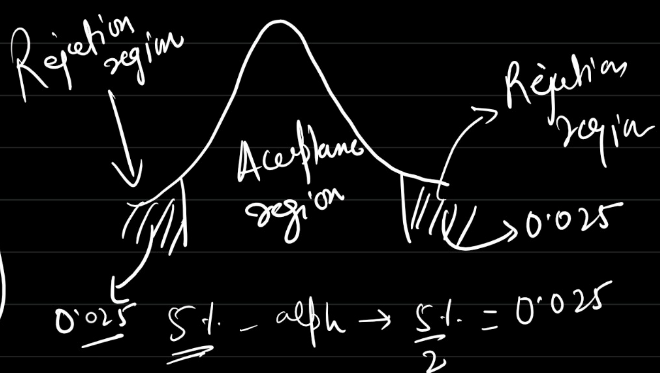


Scen-2

two tailed test

$H_0 : \mu_x = \mu_0$

$H_A : \mu_x \neq \mu_0$

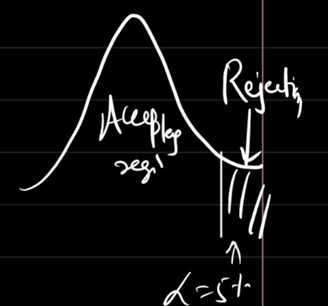


Scen-3

One tailed test (Right tail)

$H_0 : \mu_x = \mu_0$

$H_1 : \mu_x > \mu_0$



Statistical tests

1. Z test] → Average
2. t test]
3. Chi-square test → Categorical data
4. ANOVA test → variance.
5. F-test.

Q. Suppose a child Psychologist says that the average time working mother spend talking to their children is upto 11 minutes per day.

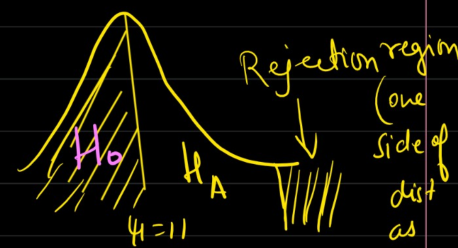
To test the hypothesis, you conducted an experiment with random sample of 100 working mother and find that they spend 11.5 minutes per daily talking with their children. Assume prior research suggest that the population std deviation is 2.3 mins.

Conduct the test with 5% level of significance. ($\alpha = 0.05$)

① **Step-1** frame the hypothesis.

$H_0 : \mu \leq 11$

$H_A : \mu > 11$



② level of significance = 5% , 0.05, one tail test

③ type of test

Since S.S ≥ 30 & σ_{pop} given, You will use Z test

$$Z_{score\ of\ \bar{x}} \,\Big|\, Z_{statistics} \,\Big|\, \text{test statist} = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$$= \frac{11.5 - 11}{2.3 / \sqrt{100}} = \boxed{2.17}$$

④ $Z_{critical}$ (test critical) / p-value.

$Z_{critical}$

$\alpha = 5\% = 0.05$

$\hookrightarrow Z_{critical\ \alpha = 0.05} = -1.64$

S.ND

Step-5 $2.17 > 1.64$ ($Z_{score} > Z_{crit}$)

$\hookrightarrow$ rejection region

reject the H_0

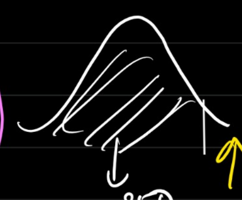
p-value approach.

$\alpha = 0.05$

$Z_{statics} = 2.17$

$Z_{statics} \rightsquigarrow$ p-value.

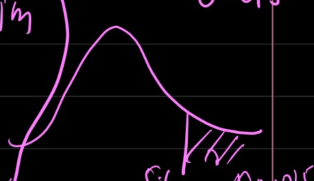
P_{value} correspond to $2.17 = 0.9850$



$= 1 - 0.9850$

$\Downarrow$

0.015



pvalue < 0.05

reject the H_0

Conclusion : the working mother spend >11 mins. Step 5

>11 mins

Q. Sachin made a statement that Avg Age of Deloitte employees is 45 years.

Agrim took a sample of 100 employee of deloitte and observed the mean value of sample came out to be 49 years. Assuming population std dev from prior research is 3 years.

test the hypothesis with 5% level of significance.

→ Step-1 → $H_0 : \mu = 45$

$H_A : \mu \neq 45$

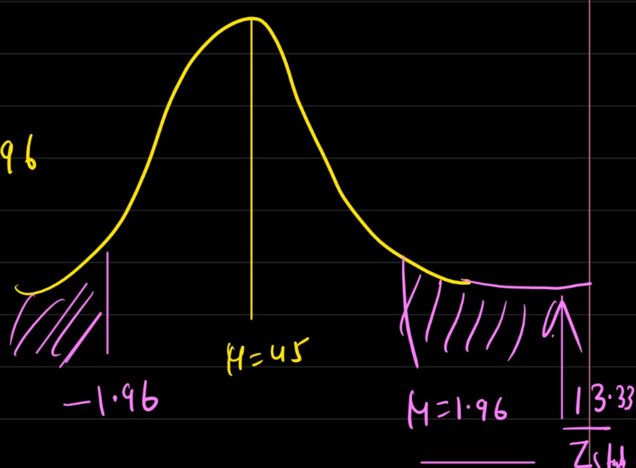
Step-2 – level of significance $= \frac{0.05}{2} = 0.025$ (why? – because its two tail test. ⇓ in H_A $\neq$)

Step-3 type of test = Z_{test}

$$Z_{\text{statistic of } \bar{x}} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{49-45}{3/\sqrt{100}} = 13.33$$

Step-4 $Z_{critical}$

$$Z_{critical\ \alpha = 0.025} = -1.96$$



Steps. test stat $>$ Z critical — reject the H_0

$Z_{st} > Z_{cri}$

$13.33 > 1.96$ → reject the H_0

Conclusion – $\mu_{age} \neq 45$

mean

Q. $X \sim N(4, 2)$ → std

① How many std dev away 4.5 is?

② What is prob of datapoint falling below 4.5

→ ① $Z_{score} = \frac{x-\mu}{\sigma} = \frac{4.5-4}{2} = \frac{0.5}{2} = \frac{1}{4} = 0.25.$

② Calculate p-value corresponding to Zscore of 0.25 = 0.5987

the prob of a dp falling below 4.5 = 59.87%.

③ What is prob of datapoint falling above 4.5?

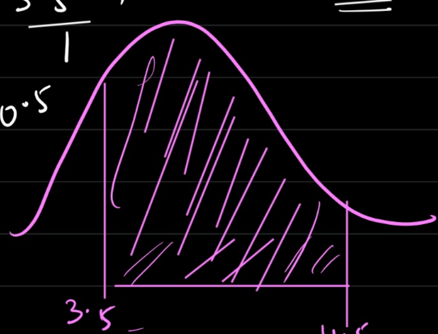
$= 1 - 0.5987 = 0.4013$

→ 40.13%

Q. What percentage of score lies between 3.5 to 4.5?? given mean = 4, std dev = 1

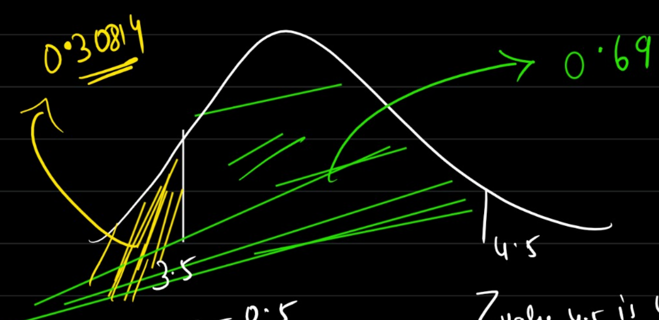
→ Zscore of 3.5 $= \frac{x-\mu}{\sigma} = \frac{3.5-4}{1} = -0.5$

Zscore of 4.5 $= \frac{4.5-4}{1} = 0.5$



Area between 3.5 & 4.5

Zvalue = −0.5 and Zvalue = 0.5

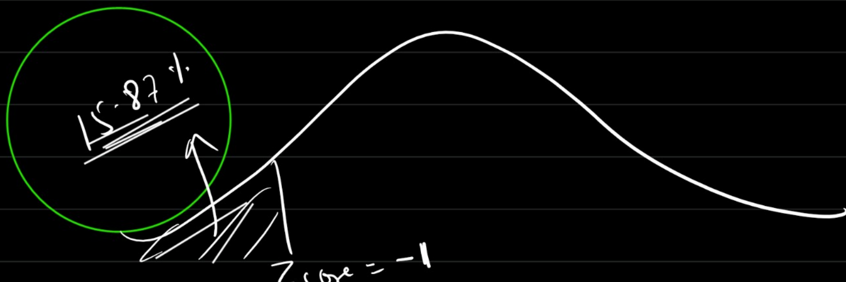


Zvalue 3.5 = −0.5   Zvalue 4.5 is 0.5

Area b/w = 0.69 − 0.30814
= 0.39
= 39%

Q. In India the avg IQ is 100 with a $\sigma = 15$. What is %age of population to have an IQ lower than 85 ?

$\rightarrow$ $\mu = 100$, $\sigma = 15$, $Z_{score} = \frac{85-100}{15} = -1$



15.86 %

Q. $\rightarrow$ IQ ≥ 85

$= 100 - 15.86$

$= 84\%$

Q. The average height of all resident in a city is 168cm and $\sigma = 3.9$.
36 individual were taken and height to be 169.5cm.
Test the hypothesis with 5% alpha.

$\rightarrow$ Step-1 $H_0 : \mu = 168$cm $\quad H_A = \mu \neq 168$cm

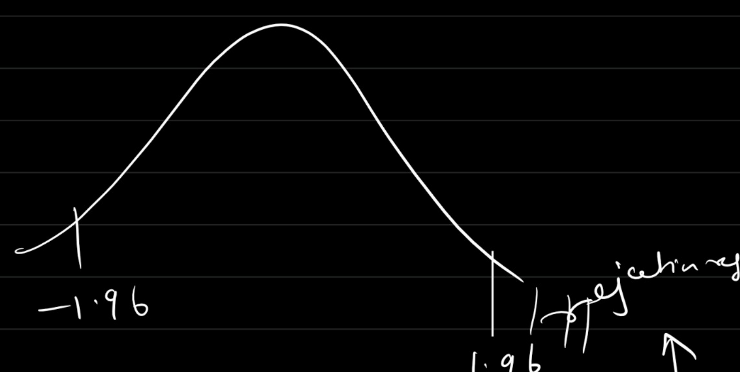
Step-2 $\alpha = 0.025$ (as its two tail tes)

Step-3 $Z_{statistics} = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{169.5 - 168}{3.9/\sqrt{36}} = 2.30$

Step-4 $Z_{critical_{\alpha = 0.025}} = -1.96$

Step 5
Conclus $= 2.31 > 1.96$

reject the H_0



$\alpha = 0.05$

$C.I = 1 - 0.05$
$= 0.95$

* in terms of pvalue

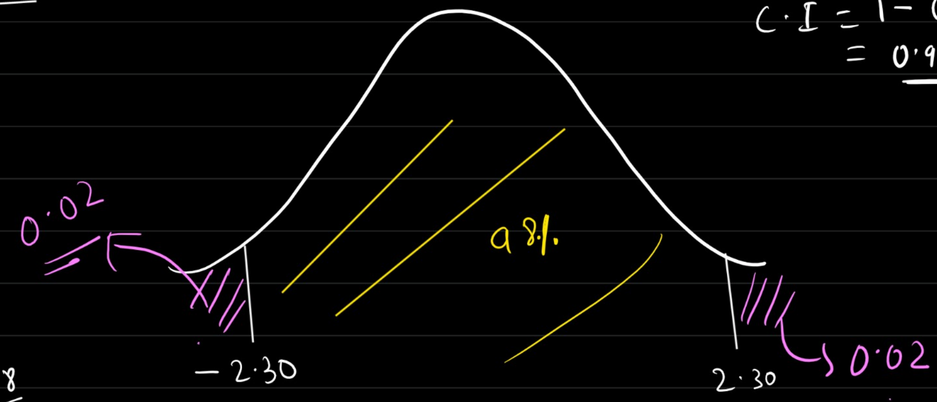
$Z_{stats} = 2.30$

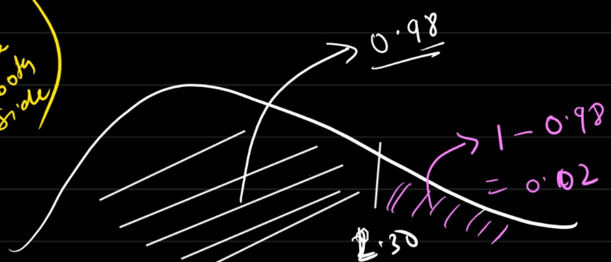
↓

pvalue.

α
↓
2

pvalue
↓
(add p value from both side)





pvalue < 0.05
↓
for two tail test
↓
pvalue $= 0.02 + 0.02$
$= 0.04$

$(0.01 \approx 0.02)$

$0.0208 < 0.05$

Reject the H_0

Q. A factory manufactures the bulb with an average warranty of 5 years with $\sigma = 0.50$

A bulb malfunctions in less than 5 years is a claim given by a worker.

To test the claim a sample of 40 bulbs were taken and avg time of bulb is 4.8 years.

2% of significance level test the hypothesis.

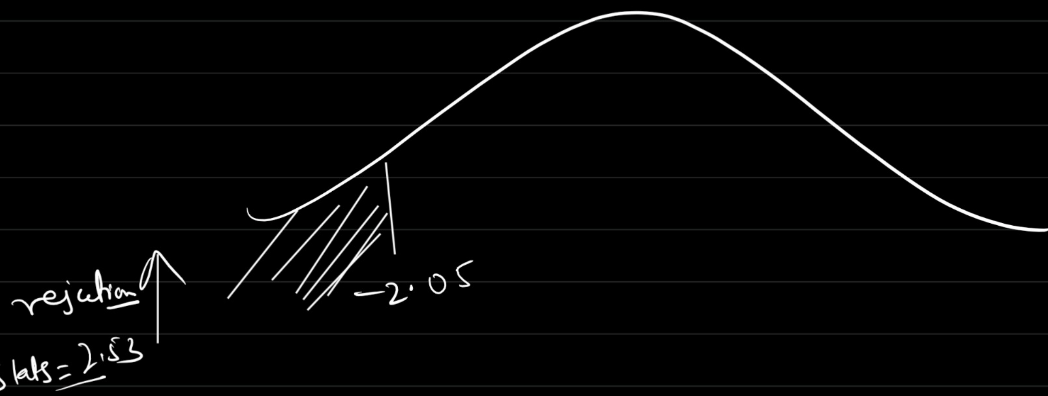
→

Step-1 → $H_0: \mu = 5$, $H_A: \mu < 5$

Step-2 $\alpha = 2\%$ — 1 tail test

Step3 $Z_{score} = \dfrac{4.8 - 5}{0.5/\sqrt{40}} = -2.53$

Step - 4 $\quad Z_{critical}\ \alpha = 0.02 = -2.05$



Step - 5 — You reject the H_0