How large n should be ??

$\downarrow$

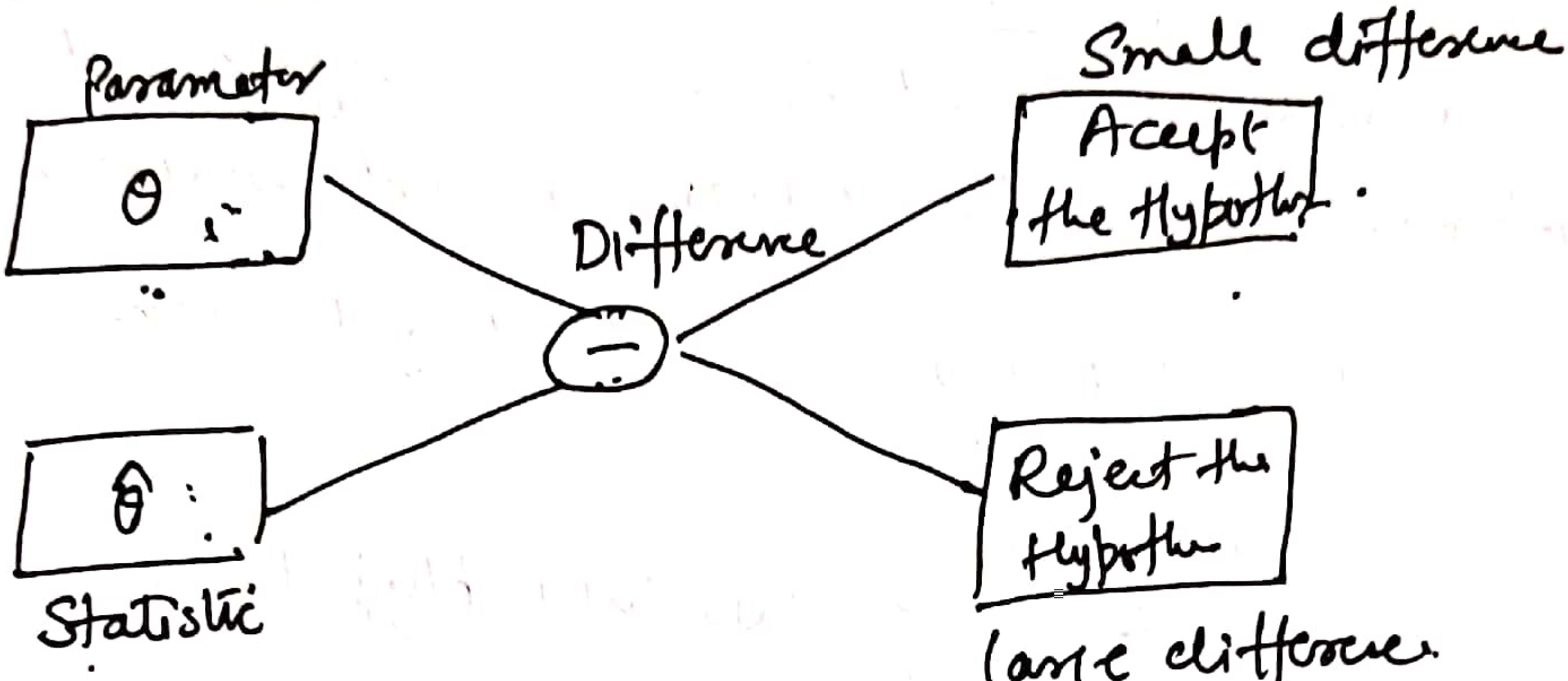
Sample size.

- $\rightarrow$ For $n \geq \underline{30}$, the CLT can be applied
- $\rightarrow$ For $n < 30$, the CLT cannot be applied.

Tests ~~for~~ of Hypothesis :

Generally, some information regarding the feature of the population may be available to us and we may like to know whether the information is tenable in light of the random sample drawn from the population. This type of problem is known as Tests of Hypotheses problem.

* HT is the process of making educated guesses about a population based on a sample drawn.
* HT involves making guesses about the difference of the parameter and the statistic.



Hypotheses: A premise / claim that we want to test / investigate / validate.

1) **Null Hypotheses:** A hypotheses to be tested is called the null hypotheses. It is denoted by H_0

→ There is no difference between the parameter and statistic.

eg Find the null hypotheses s.t the 3^{rd} year students of IITJ had a mean ~~IITJ~~ JEE score of 400.

$H_0 : \mu = 400.$ → value to be tested

Population Mean

⇓

$\boxed{H_0 : \mu - 400 = 0}$

Alternative Hypotheses: An hypotheses which compliments the Null Hypotheses is called the Alternative Hypotheses. It is usually represented by H_1 / Ha

→ AH usualy says that there is a significant difference betwen the parameter and statistic.

eg $H_1 : \mu \neq 400$ $H_0 : \mu = 400$

$\mu > 400$ & $\mu < 400$ } Two tail Hypothesis.

$H_1 : \mu > 400$
$H_a : \mu < 400$ } ⇒ One-tail Hypotheses

eg A pharma company manufactured a particular medicine which supposed to have 14 ml of active ingredient. What are the null and alternative hypotheses?

$H_0 : \mu = 14$ → Population has a mean active ingredient equal to 14 ml.

$H_1 : \mu \neq 14$

Eg The IITJ director wants to test if it is true what faculty says - B.Tech students use the laptop on average of 4 to 5 hrs a day. (4-4)

⇒ $H_0: \mu \in [4, 5] \equiv 4 \leq \mu \leq 5$

$H_1: \mu \notin [4, 5]$

* During, the hypotheses testing, a statistician must decide on how much evidence is necessary to Accept / Reject the hypotheses.

1:39

<u>Level of Significance</u>: LOS is the probability that a the statistic lies in the critical region / Region of Rejection.

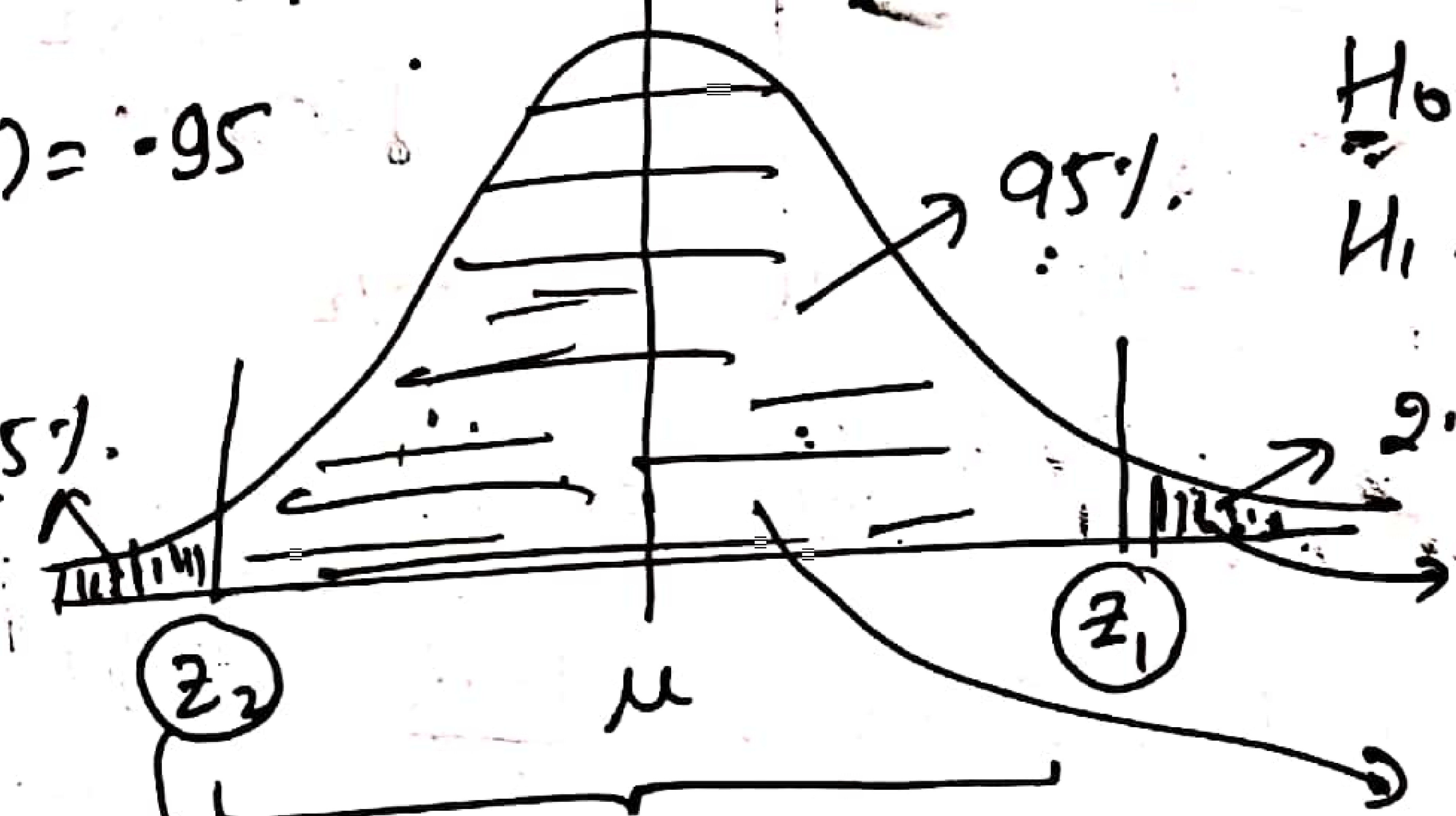
<u>Level of Confidence</u>: LOC is the probability that the statistic lies in the Region of acceptance.

Consider a large sample. Then the underlying sampling distribution is normal distribution. If I say that to accept the Hypothesis in 95%..

$P(z_2 < X < z_1) = .95$

$H_0 : \mu = \mu'$.

$H_1 : \mu \neq \mu'$.



Region of Rejection

Region of acceptance

Two-tail Analysis

⇓

Two-tail Hypothesis problem

Confidense Interval

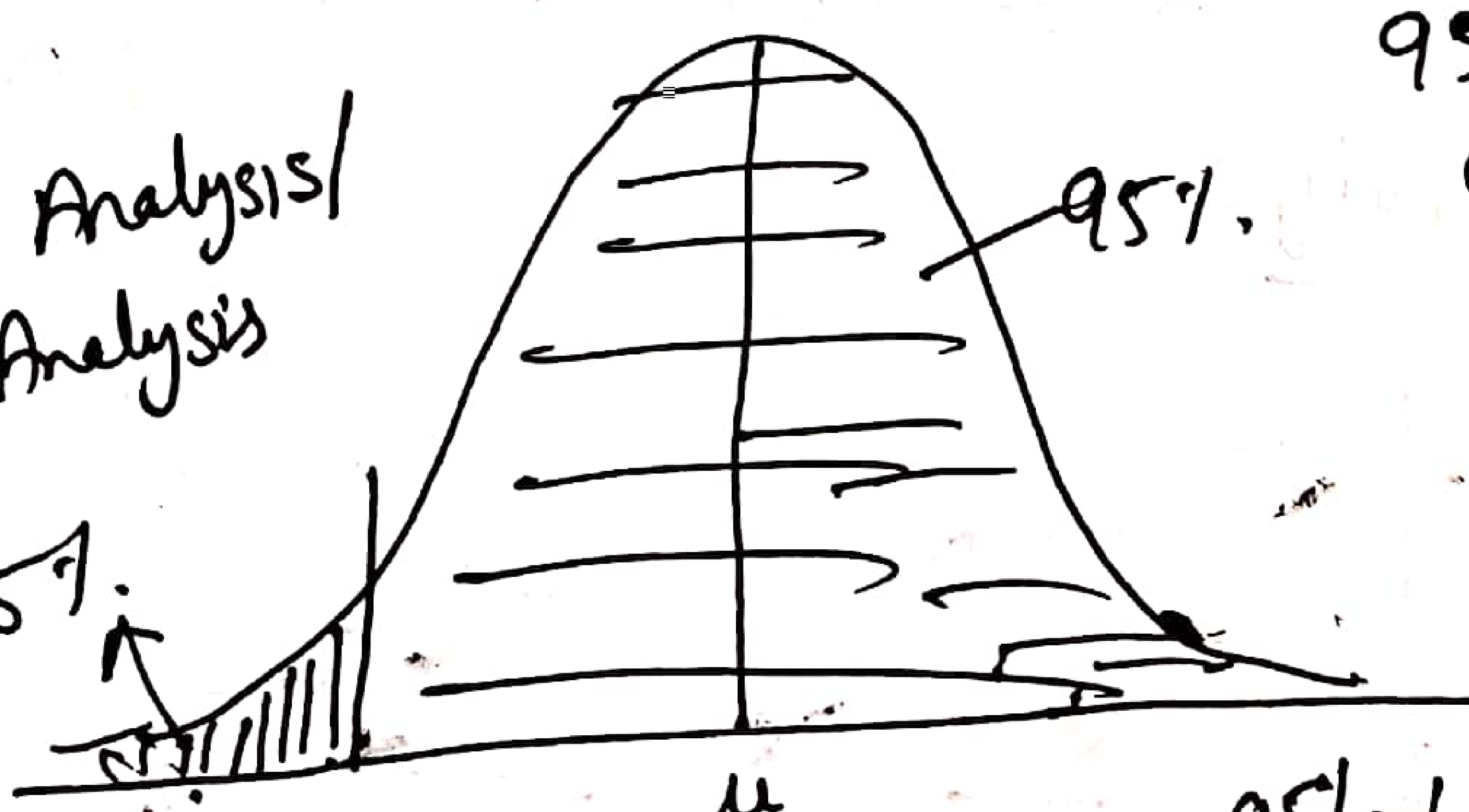
Critical values.

95% level of Confidence

Single tail Analysis / left sided Analysis

⇓

Single tail Hypotheses problem



Right tail analysis

⇓

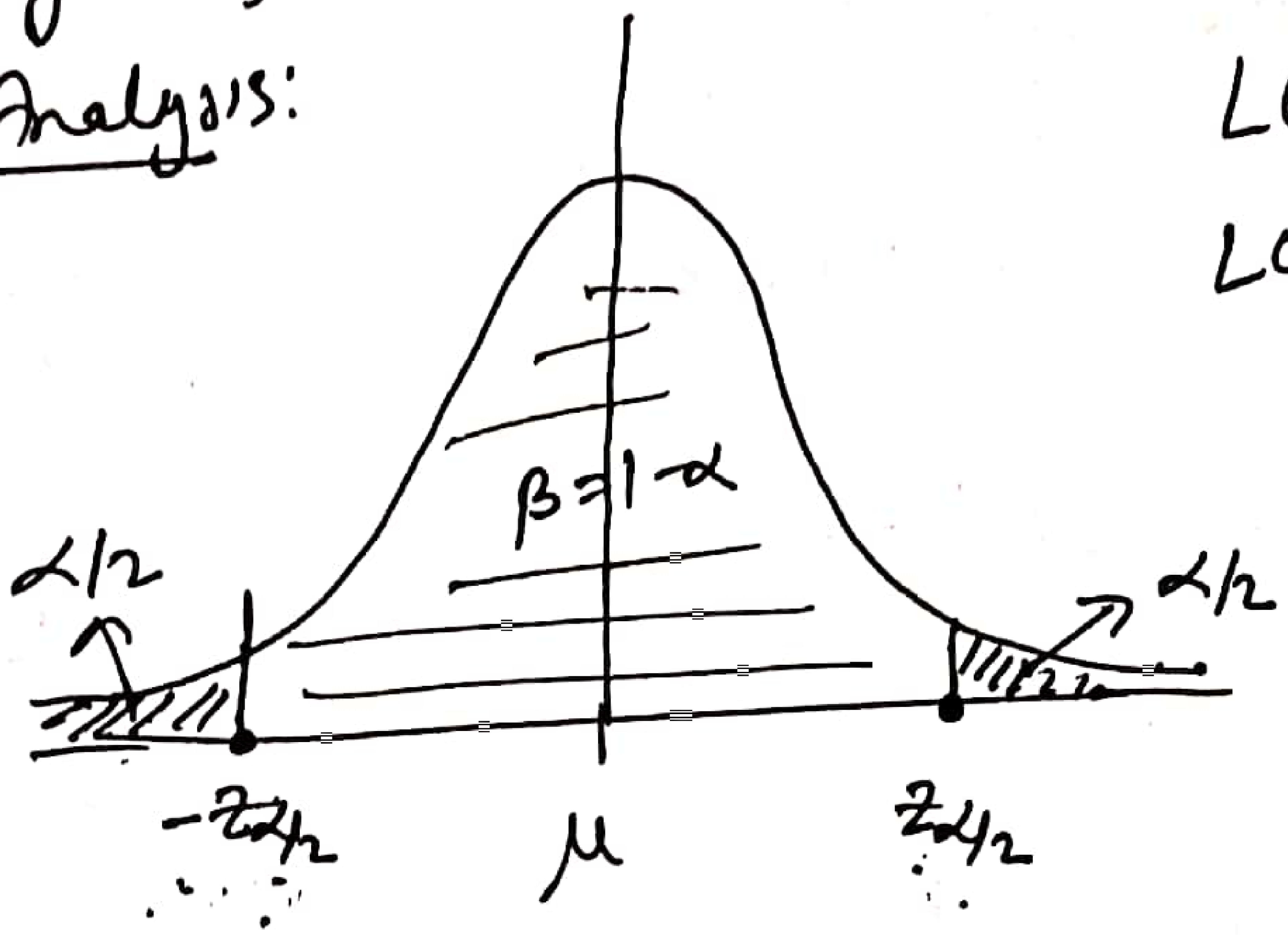
Single tail Hypotheses Problem.

Critical Points: Points, which acts as the boundary of the Region of acceptance and the Region of Rejection are called Critical Points.

Two tail Analysis:

LOS (α)

LOC ($\beta = 1-\alpha$)



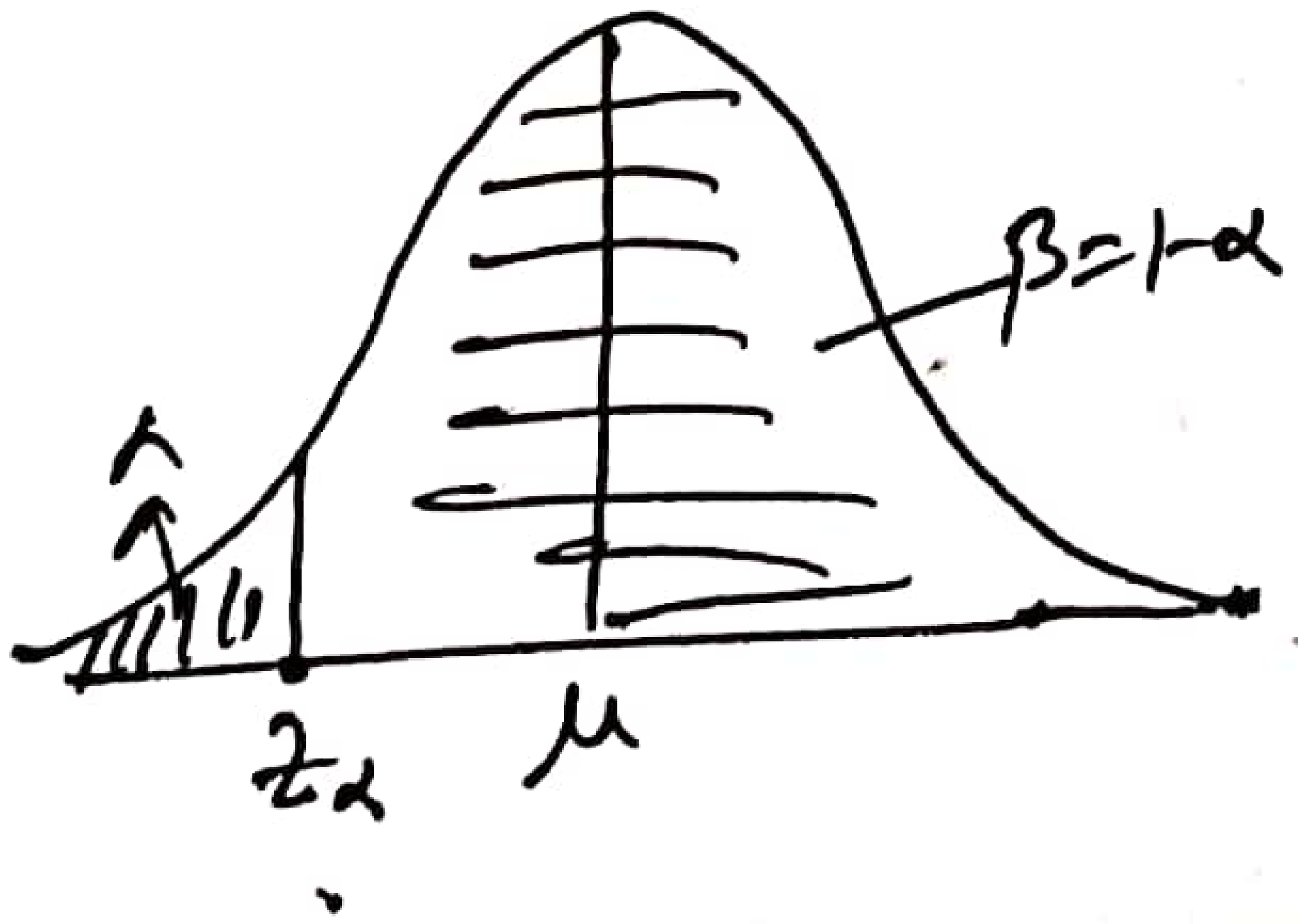
$$P(-z_{\alpha/2} \le X \le z_{\alpha/2}) = 1-\alpha$$

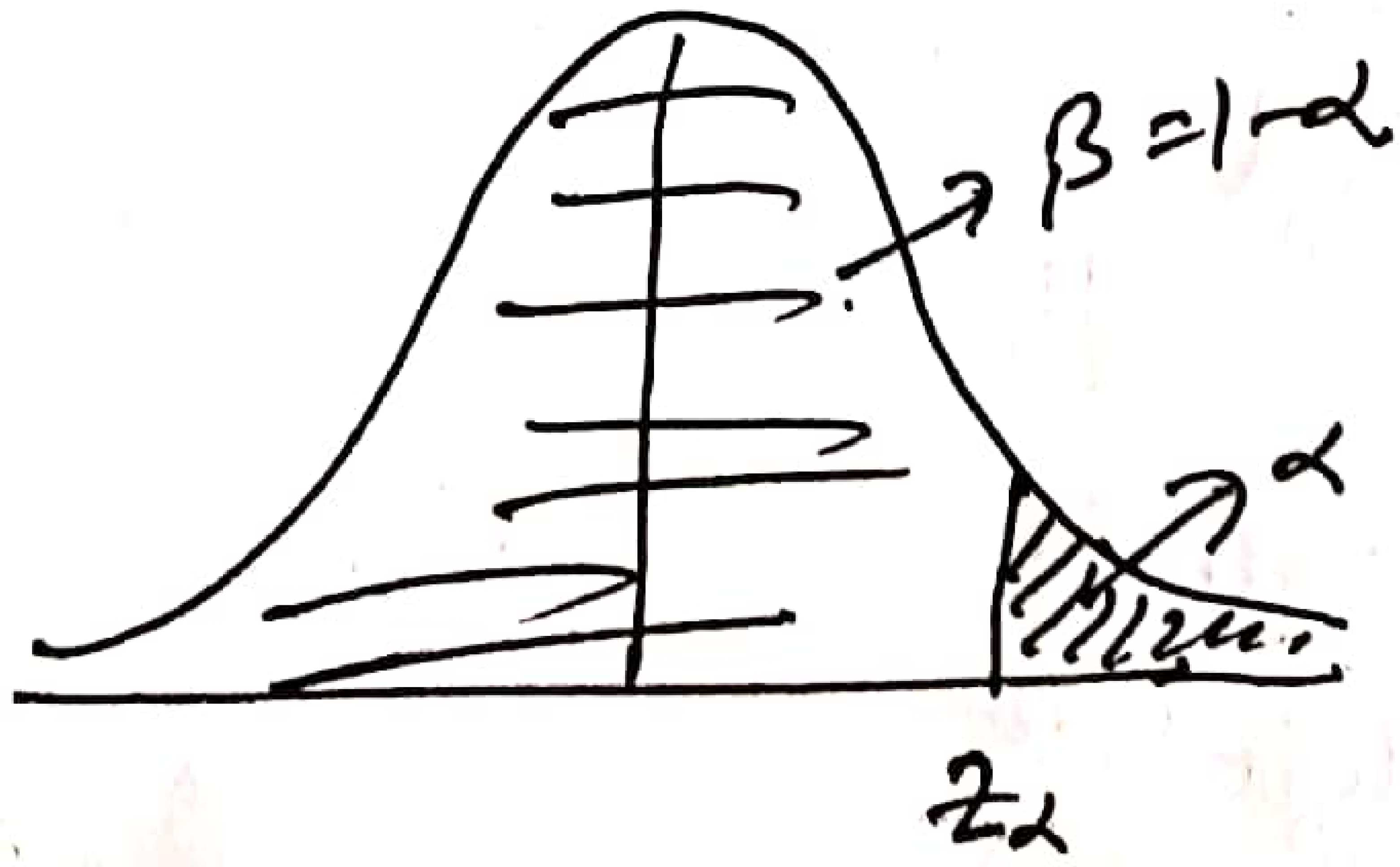
$$\Rightarrow P(|X| \le z_{\alpha/2}) = \underbrace{1-\alpha}_{\beta}$$ // with respect to level of confidence

$$\Rightarrow 1 - P(|X| \ge z_{\alpha/2}) = 1-\alpha$$

$$\Rightarrow P(|X| \ge z_{\alpha/2}) = \alpha$$ // with respect to Level of Significance

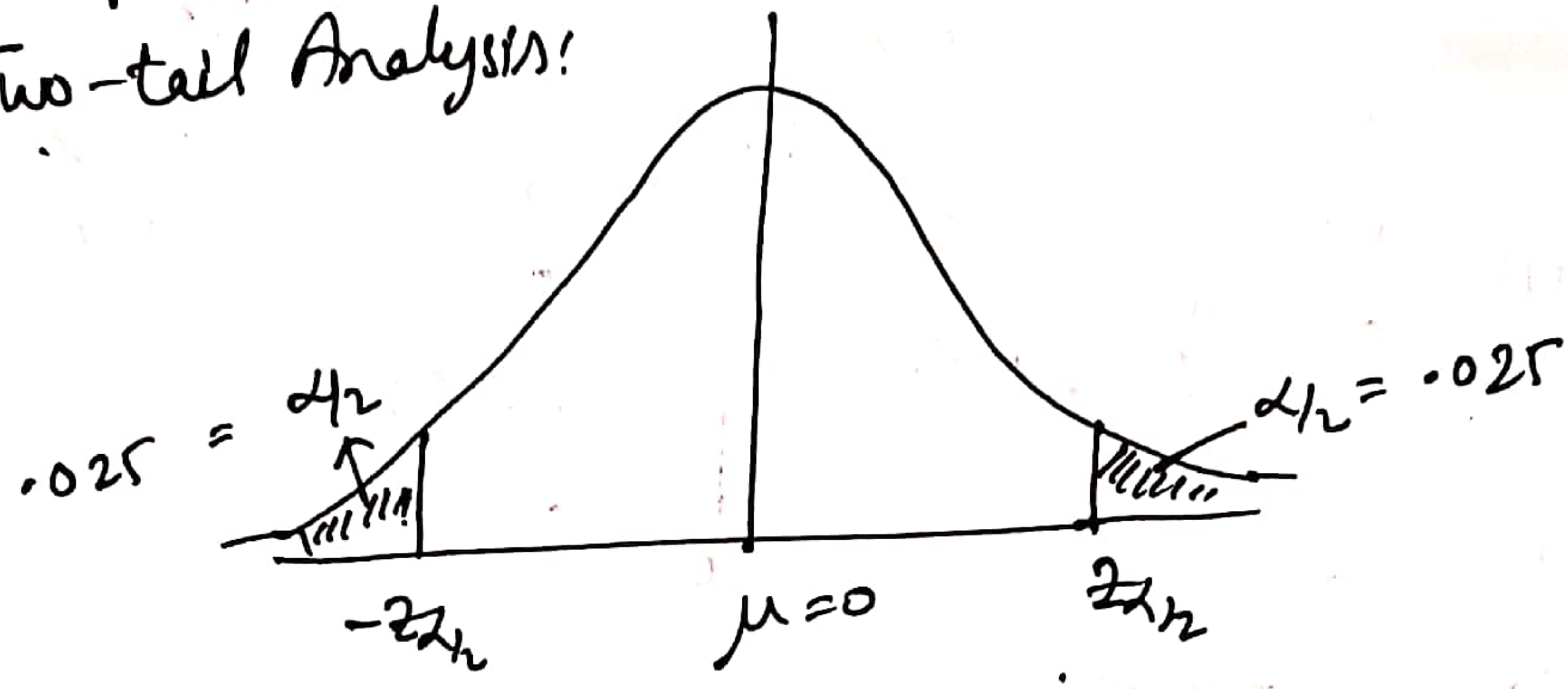
One tail Analysis:





Q Find the critical values for single and two-tail analysis with a 0.05 significance level. Assume that you have a fairly large sample.

(i) Two-tail Analysis:



$$P(X > z_{\alpha/2}) = 0.025$$

$$\Rightarrow 1 - P(X < z_{\alpha/2}) = 0.025$$

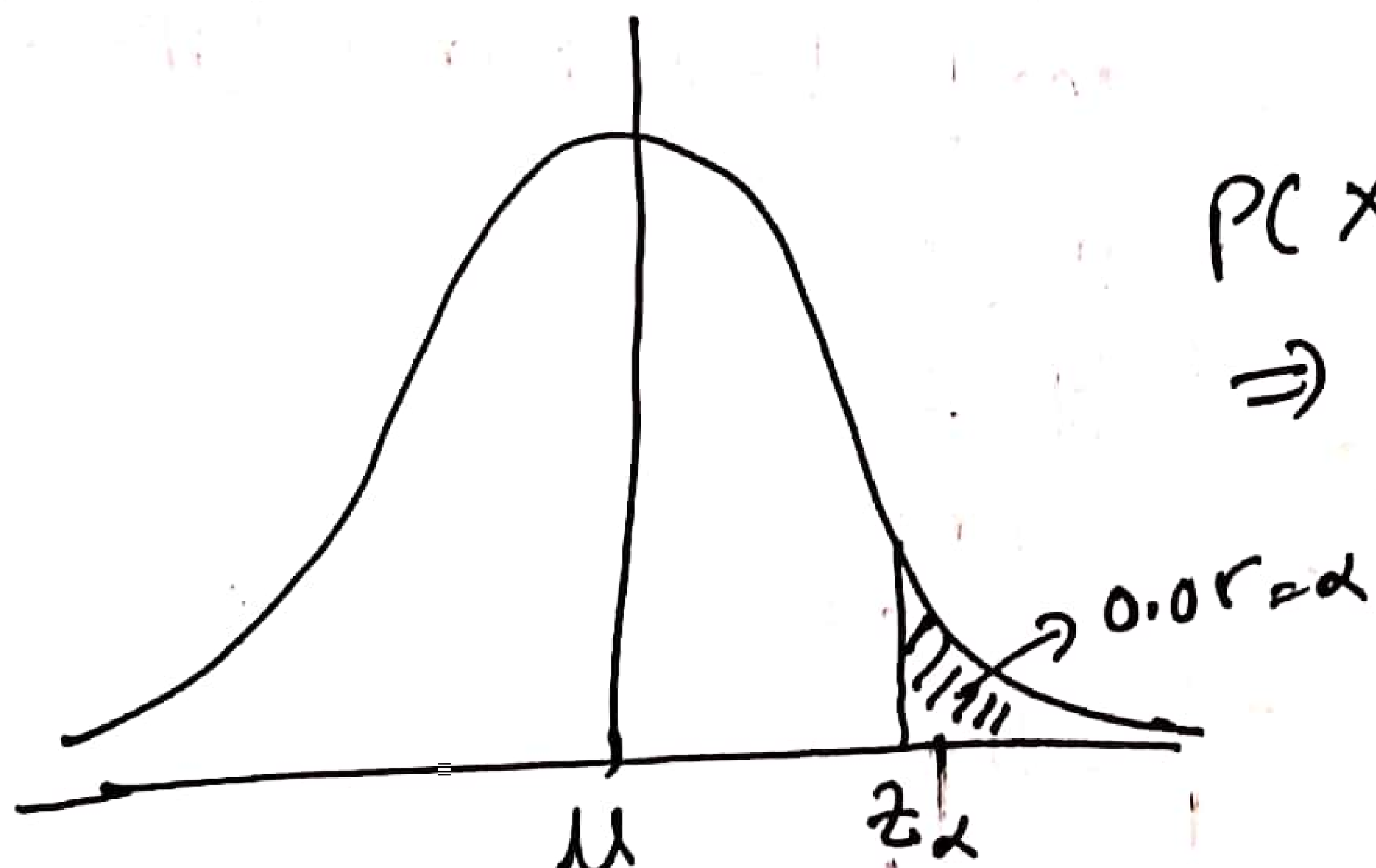
$$\Rightarrow P(X < z_{\alpha/2}) = 1 - 0.025 = \beta + \alpha/2$$

$$\Rightarrow z_{\alpha/2} = 1.96$$

$$\Rightarrow -z_{\alpha/2} = -1.96$$

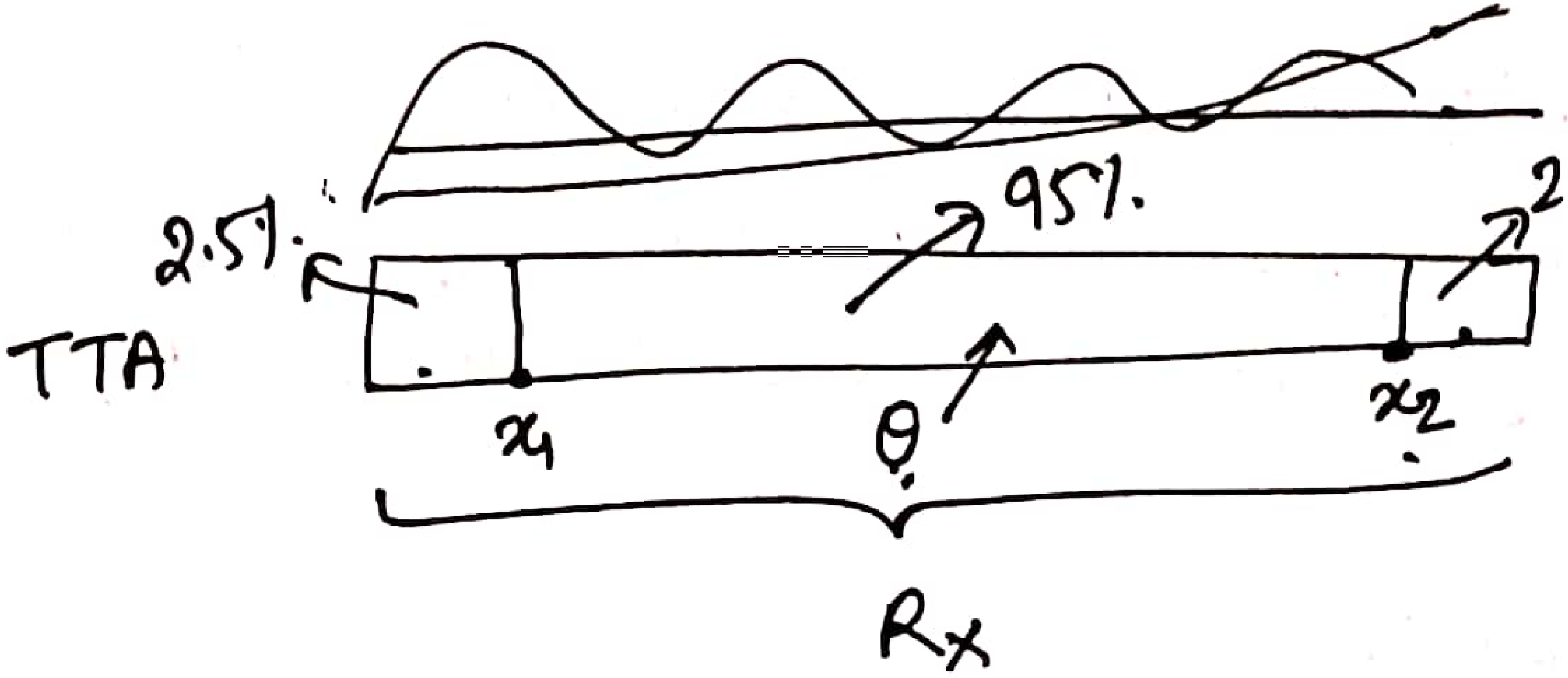
} H.W

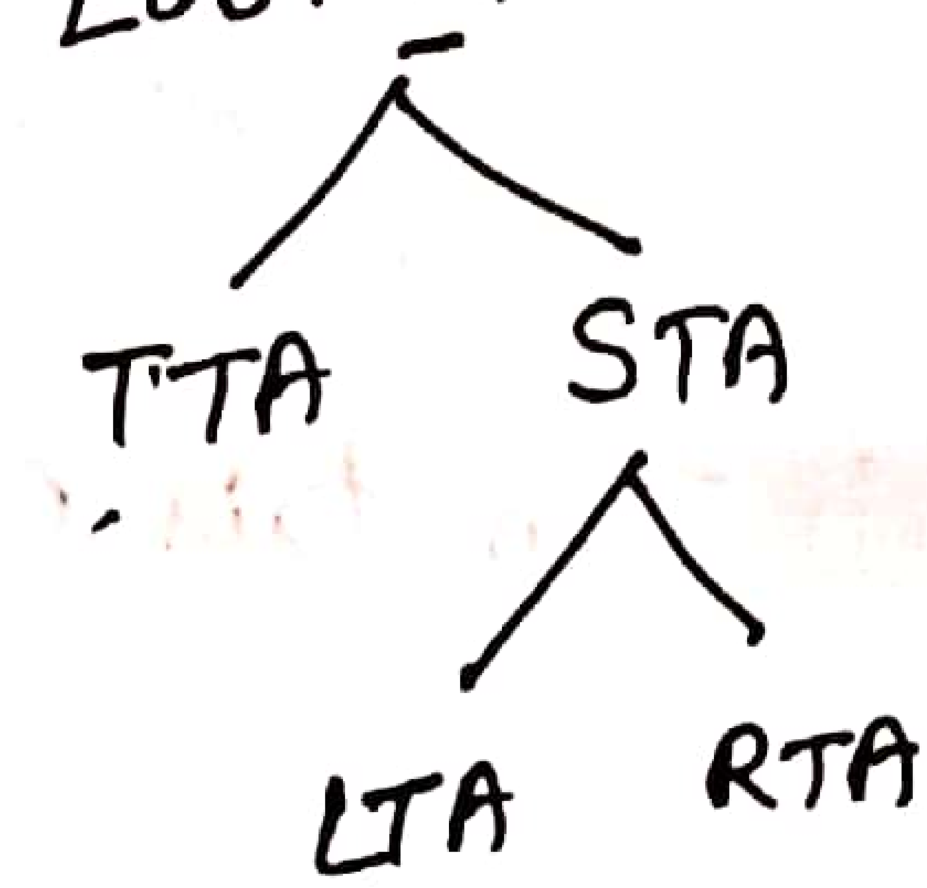
(ii)

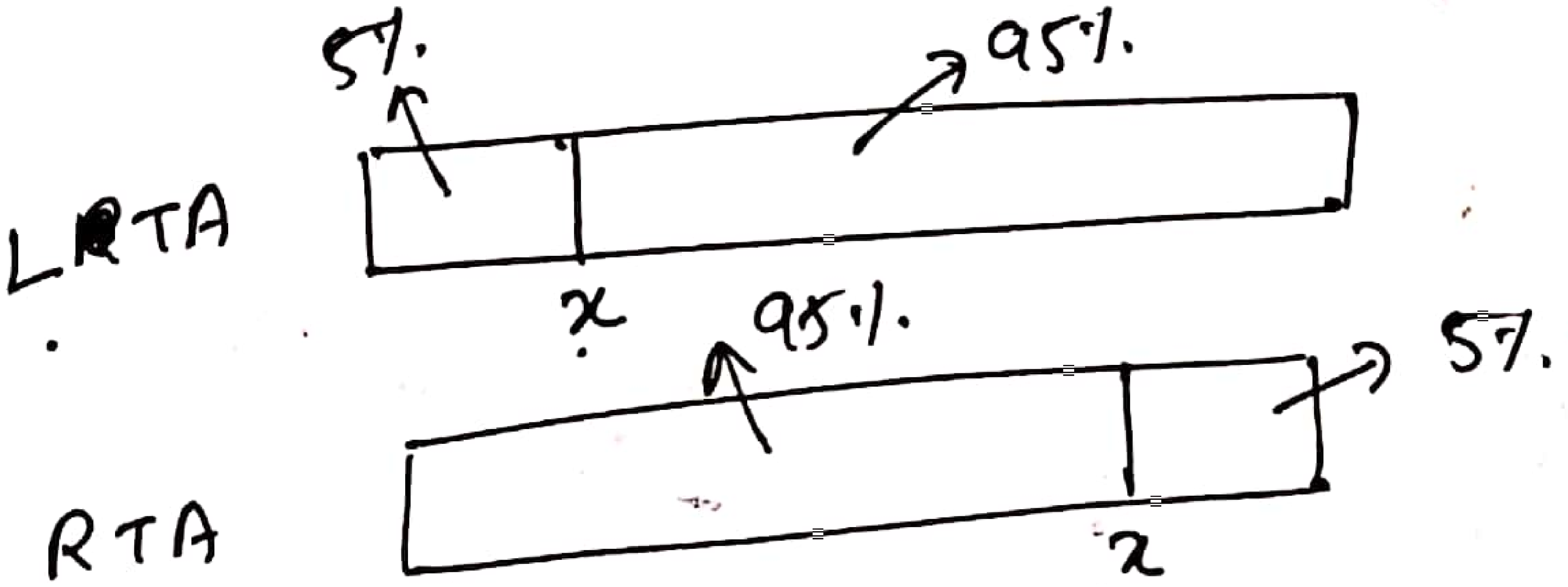


$$P(X > z_\alpha) = 0.05$$

$$\Rightarrow z_\alpha = 1.65 \quad \text{(HW)}$$







Errors in HT/T of H:

Let us consider H_0 and H_1 be the Null and Alternative hypotheses corresponding to some problem. Then we have following four cases:

(i) A true Hypothesis is rejected → error (Type 1)

(ii) A " " " accepted } Correct Decision

(iii) A false " " rejected } Correct Decision

(iv) A false " " accepted → error (Type II)

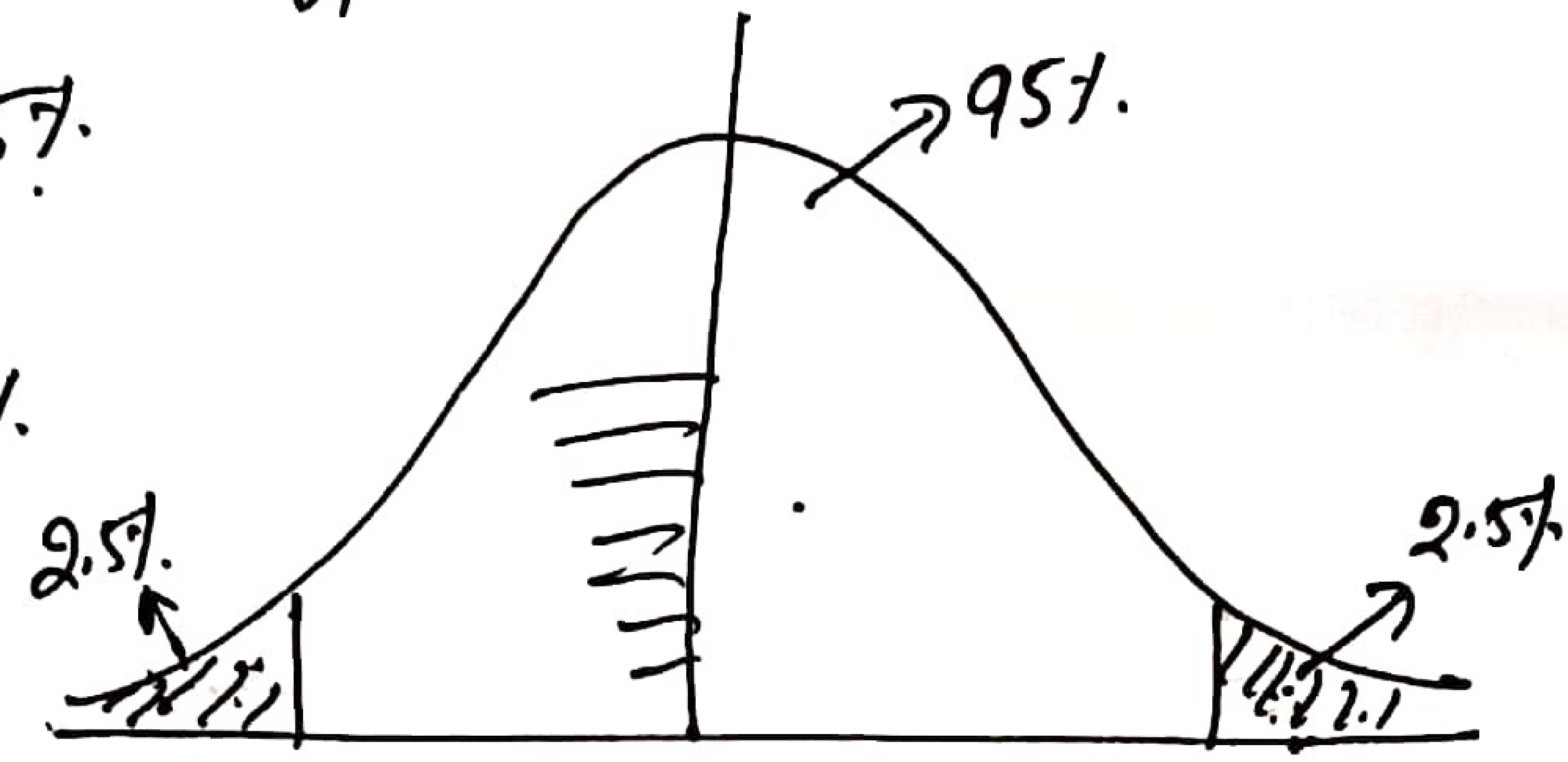
	$H_0 \to T$	$H_1 \to T$
Reject H_1 / Accept H_0 ✓	CD	Type II error
Accept H_1 / Reject H_0	Type I error	CD

Type 1 error:

Rejection of Null Hypotheses when it is true

eg LOS = 0.05 ≈ 5%.

∴ LOC = 1 − 0.05
= 0.95 ≈ 95%.



→ Type I error can be minimized by taking smaller LOS.
⇒ α should be ASAP.

Type I error (LOS = 5%)
> Type I error (LOS = 2%)
with respect to same problem.

Type II error:

Reject // Cannot accept

Reject a true Alternative Hypotheses
OR
Accepting Null Hypotheses when it is not true

→ In order to minimize Type II error, we need to follow power of test procedure.

→ (1) Producer / Manufacturer — Type I
(2) User / Consumer — Type II