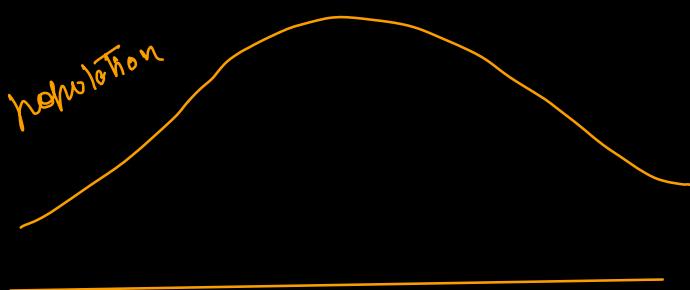


## Inferential Stats

1. Central limit theorem.
2. Type 1 - Type 2 errors
3. One tail vs Two tail
4. Z-test, t-test, p-value
5. Z-test for proportional population
6. Chi-square
7. ANOVA

## Central Limit Theorem

If you have a distribution (Gaussian or non gaussian)



If we now take multiple samples from this distribution.

$$S_1 = [x_1 \dots x_{\bar{x}}] \rightarrow \bar{x}_1$$

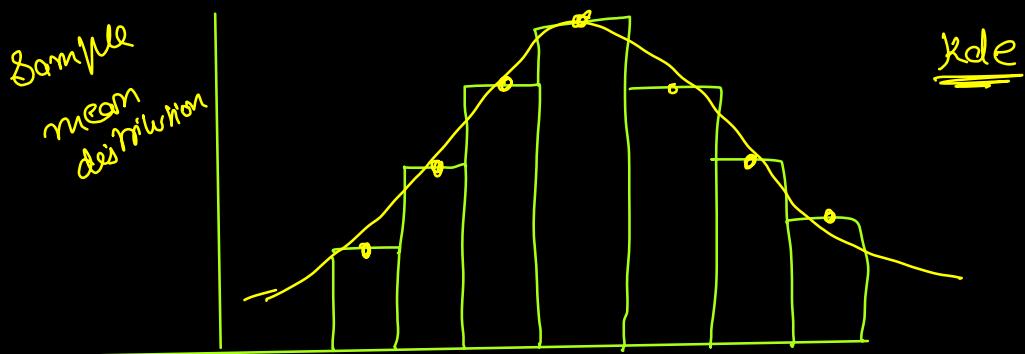
$$S_2 = [x_1 \dots x_{\bar{x}}] \rightarrow \bar{x}_2$$

$$S_3 \rightarrow \bar{x}_3$$

⋮

$$S_n \rightarrow \bar{x}_n$$

then, if we now plot all sample means  
on a graph. that will actually be  
a normal / gaussian distribution

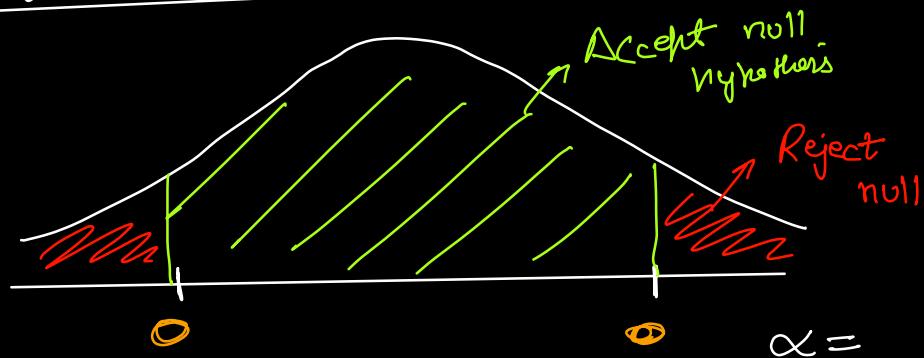


$n \geq 30$  usually applies for a  
non gaussian distribution

Works all the time.

In inferential statistics with any sample data, we will thus be focusing on normal / Gaussian distribution.

## Confidence Interval



A confidence interval from sample data is a range of values which are likely to have true parameters at some confidence level.

## Point estimation

The value of any statistic that estimate  
the value of a parameter is  
called point estimate.

$$\begin{array}{ll}\bar{x} & \mu \\ (\text{sample}) & (\text{pop}) \\ 2.5, 2.8 \\ 3.1, 3.3 & 3\end{array}$$

$\bar{x}$  is point estimate of  $\mu$

Inferential stats :-

We have some sample data  
from which we get value of  
hyp.

This point estimate can help us  
get the value of beh. parameter.

\* Estimate of hyp. = Point estimate  $\pm$   
margin of error.

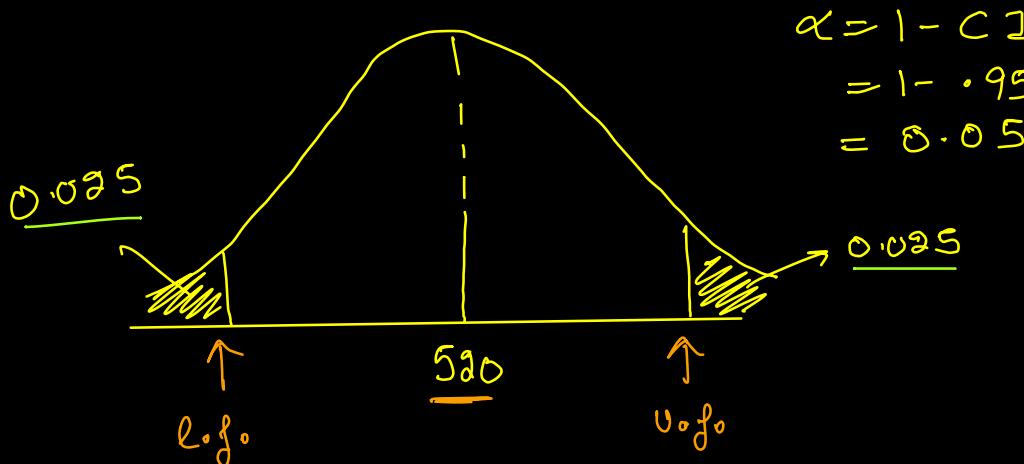
Q: On a Data science test of a company  
the  $\sigma_{\text{pop}}$  is known to be 100. [80]

A sample of 25 test takers has  
a mean of 520.

Construct a 95% CI about the  
mean?

$$\alpha = 100 \quad \bar{x} (\text{Sample}) = 520$$

$$n = 25 \quad CI = 95\% \quad \alpha = 0.05$$



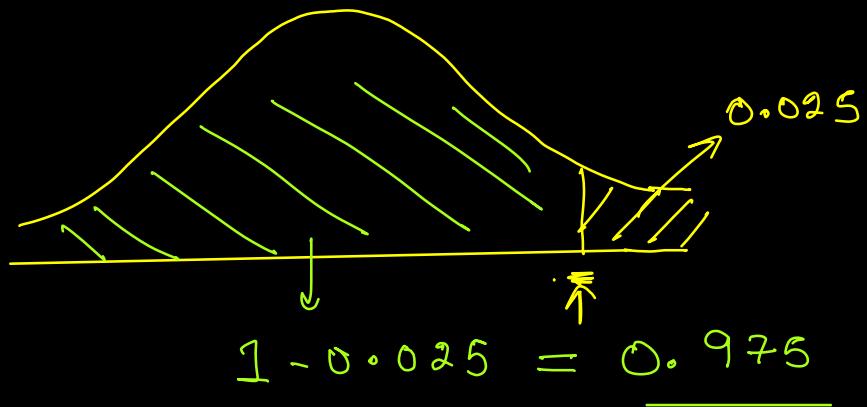
Population  $\sigma$  is given  $\Rightarrow \{z\text{-test}\}$

$$\bar{x} \pm \text{margin of error}$$

$$\text{margin of error} = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \rightarrow \begin{array}{l} \sigma \rightarrow \text{stand. deviation} \\ \sqrt{n} \rightarrow \text{sample size} \end{array}$$

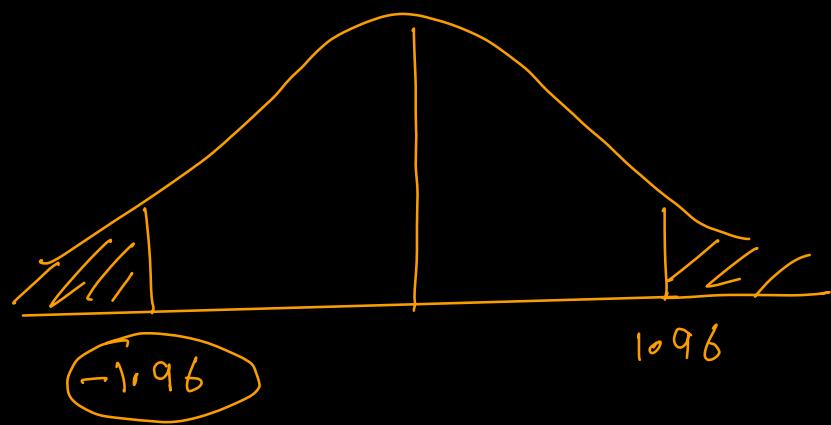
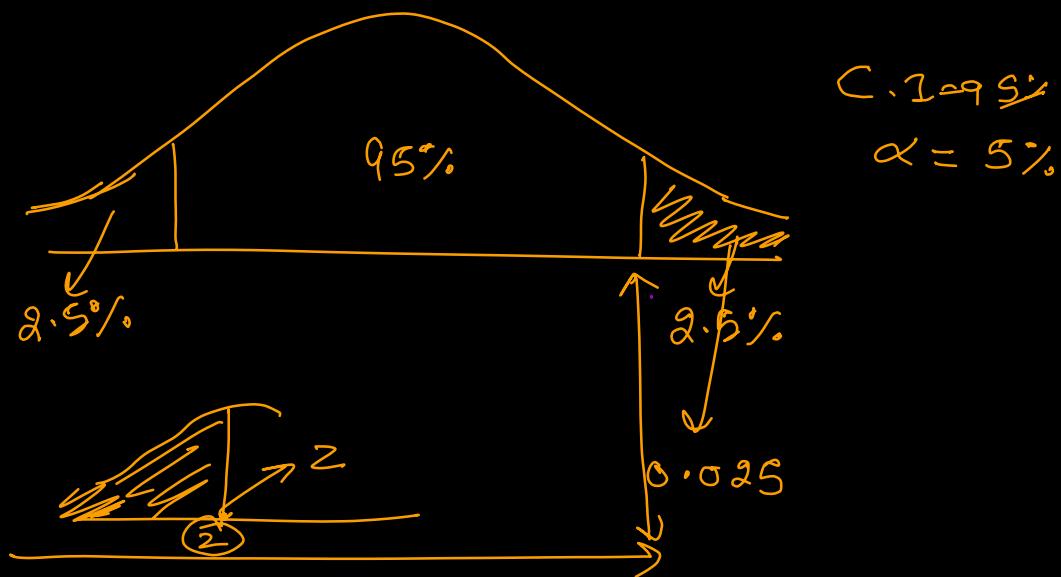
$Z$ -test is a test which is used to compare your sample mean to your population mean when population  $\sigma$  is known & sample size is large enough

$$Z_{\alpha/2} = Z_{0.5/2} = \underline{Z_{0.025}}$$



$$\text{body area} = \underline{0.975}$$

$$\text{tail area} = 0.025$$



margin of error

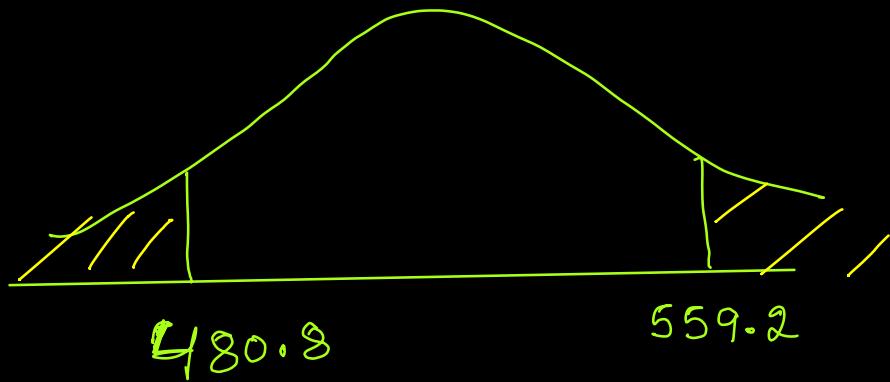
$$= 1.96 \times \frac{100}{\sqrt{25}} = 2 \times \frac{\alpha}{\sqrt{n}}$$

$$= 39.2$$

population estimate = 520 ± 39.2

$$\text{L.F} = 520 - 39.2 = 480.8$$

$$\text{U.F} = 520 + 39.2 = 559.2$$

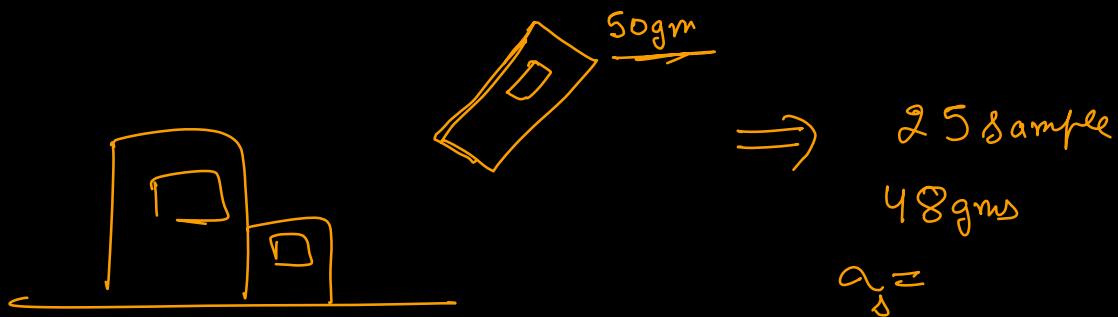


∴ we are 95% confident that the population mean is b/w 480.8 & 559.2.

A standard deviation = 80

## T - test

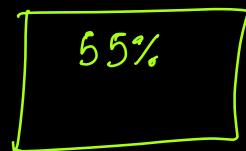
t - test is used to compare mean of 2 groups. Normally when sample size is less and standard deviation.



Drug A



Drug B



## Z - test

- ①  $n \geq 30$
- ② population or sample  $\sigma$  is given

## t - test

- ① sample  $\sigma$
- ②  $n < 30$

Q.- In a data science test, a sample of 25 test takers have

a mean of 520 and  $\sigma$  of 80  
( $\sigma$ )  $\downarrow$   
sample

Construct a 95% CI about mean?

$$n = 25 \quad \bar{x} = 520 \quad \sigma_{\text{sam}} = 80$$

( $\sigma$ )

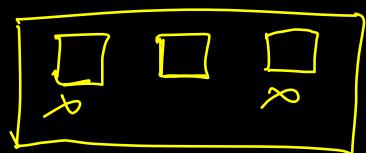
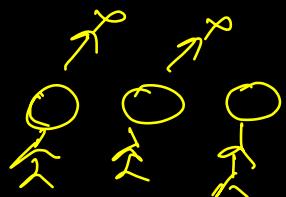
Population estimate =  $\bar{x}$  ± error



$$t\text{-test} = t_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

We need to find degree of freedom in t-test which is  $n-1$ .  $\Rightarrow$  dof

dof = how many choices we have to estimate



1  
1

$$t_{\alpha/2} = t_{0.025}$$

$$\underline{df} = n - 1 = 24 = 25 - 1$$

$$t_{0.025} = 2.064$$

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

$$520 - 2.064 \frac{80}{5}$$

$$\Rightarrow 486.97$$

We calculate  $t_{\alpha/2}$  for checking  
value in t-table

but error is  $t_{\alpha/2} \frac{s}{\sqrt{n}}$

Q:- estimate the avg. wt of sharks  
in the ocean?

## Type 1 error vs Type 2 error

$H_0$ : coin is fair (null)

$H_1$ : coin is not fair (alternate)

Experiment will be done

null true or null false

Real

Outcome

O1: we reject the null hypo.  
when in reality it is false

O2: we reject null hypo. when  
in reality it is true → type 1  
error

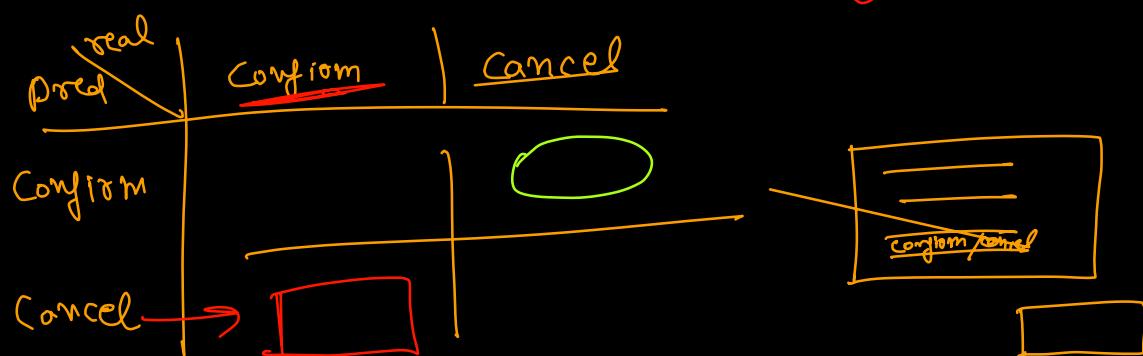
O3: we accept the null hypo. → type 2  
error.  
when in reality it is false.

O4 - we accept the null hypo.  
when in reality it is true

$H_0$

real predict	Cancer	no Cancer
Cancer	✓	✗
No Cancer	✗	✓

↓



		Real	
		+ve	-ve
Pred	+ve		Type I error
	-ve	Type II error	

$H_0 \Rightarrow$  null hypothesis, it assumes that there is no statistical diff b/w groups

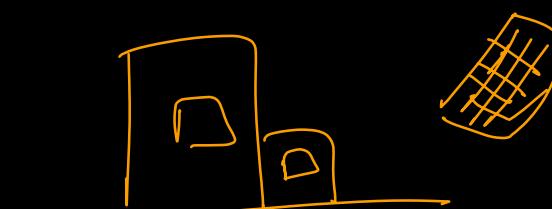
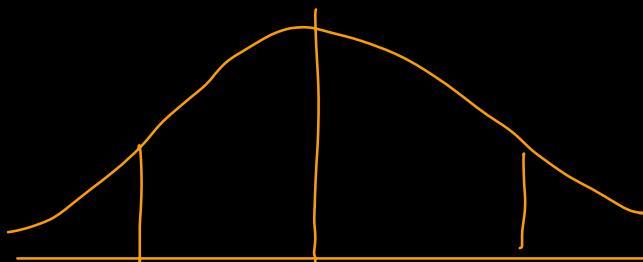
	null hyp P T	null hyp P F
Reject null hyp.	Type I error	✓
Fail to reject null	✓	Type 2 error

## One tail and 2 tail test

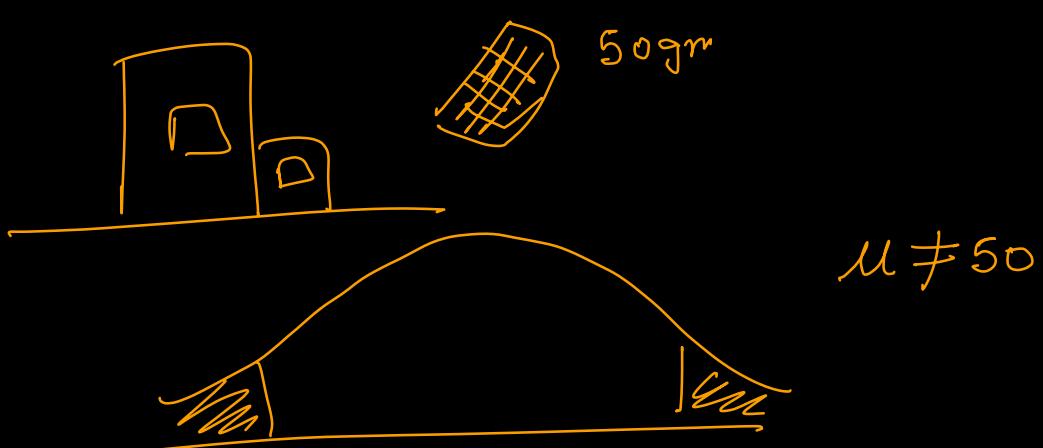
Q  $\Rightarrow$  A factory is making chocolate bars which have a mean wt of 50 gm.

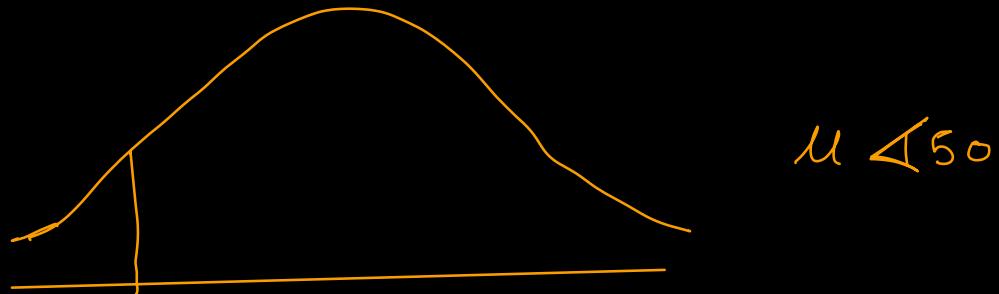
You took 25 samples & observed mean to be 49 gms. with sd. of 5.

Is the <sup>avg.</sup> wt actually 50 gm?



50 gm

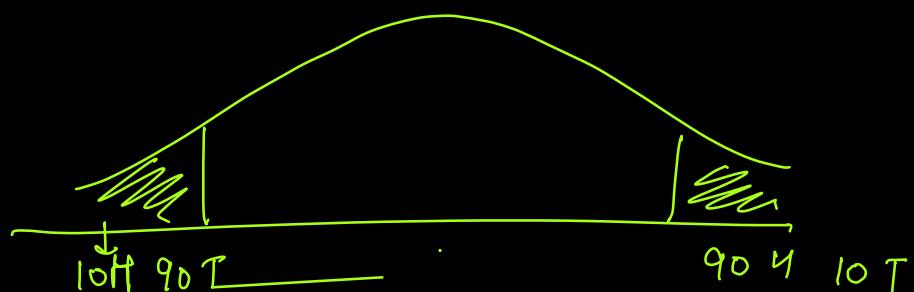
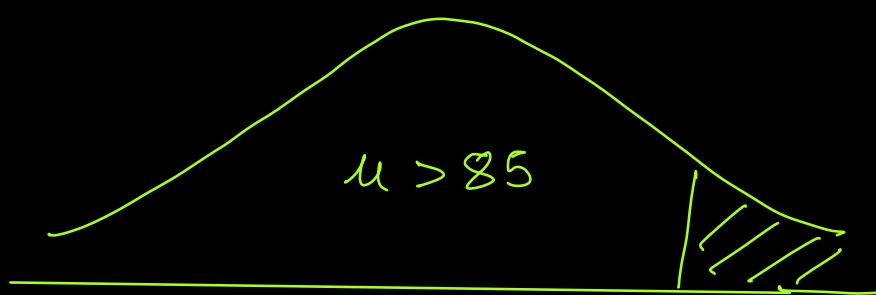
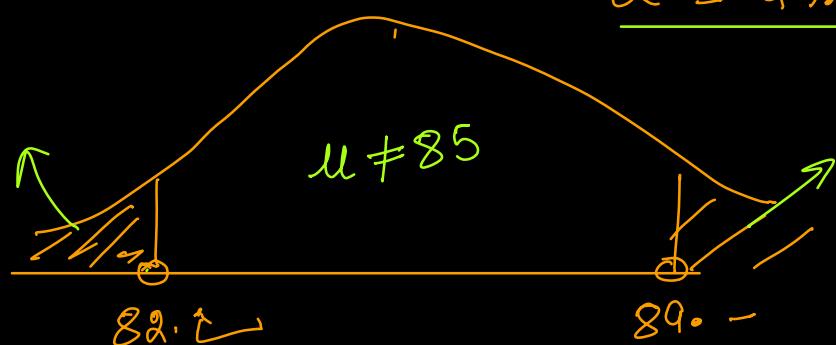




NSIT P.v = 85%

200 student sample  $P.v(\text{sample}) = 88\%$

$\alpha = 4\%$





Q  $\Rightarrow$  A factory has a m/c that fills  
80ml of medicine in a bottle

Using 40 samples, we measure the  
avg quantity dispense / filled by  
m/c to be 78ml with a s-d of 2.5.

Q  $\Rightarrow$  a) State null and alternate hypotheses

w) At 95% C.I - is there enough  
evidence that m/c is working properly?

Step 1  $H_0 \Rightarrow \mu = 80$

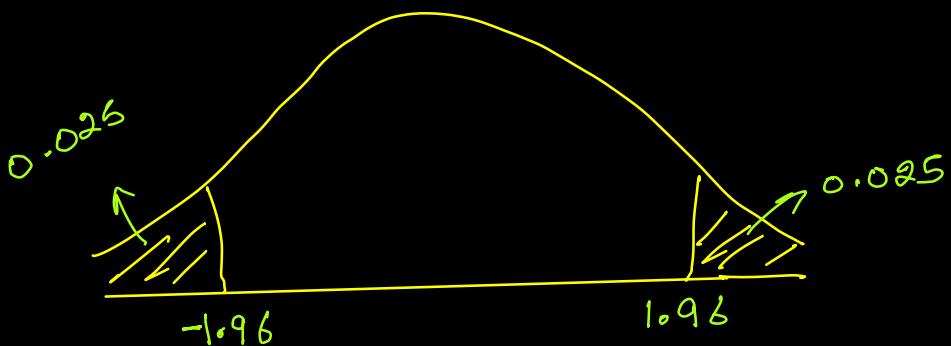
$H_1 \Rightarrow \mu \neq 80$

$n = 40 \quad \bar{x} = 78 \quad \alpha = 2.5 \quad \mu = 80$

C.I. = 95%  $\quad \alpha = 0.05$

$\Rightarrow$  z-test  
 $\Rightarrow$  2-tail

### Step 3 :- Decision boundary

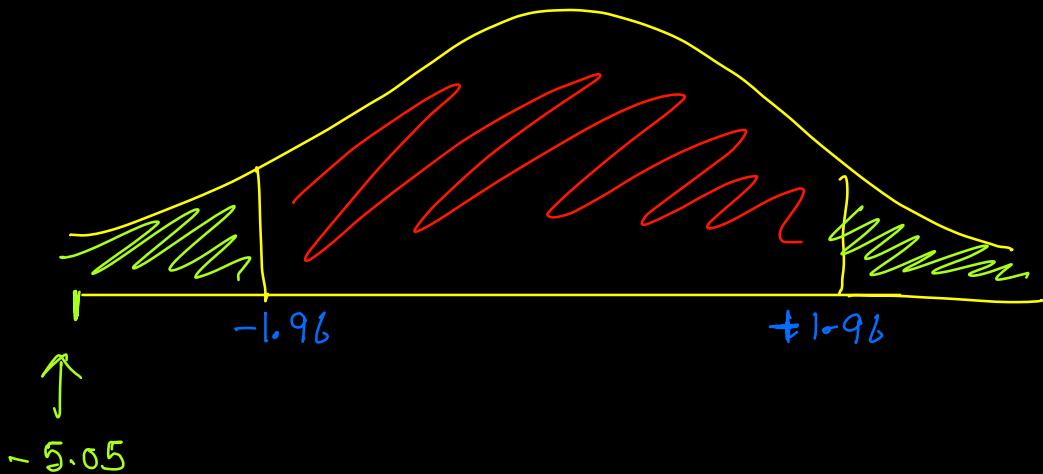


④ Calculate test statistics. for single

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

$$z = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \rightarrow \text{for sample}$$

$$= \frac{78 - 80}{2.5 / \sqrt{40}} = \boxed{-5.05}$$



⑤ Conclusion / State the result:

If  $z = -5.05$  is less than  $-1.96$   
or greater than  $+1.96$

then we reject null hyp.

with 95% confidence

$\Rightarrow$  There is a fault in the m/c.