## Hypothesis testing and Statistical Analysis

Analysis on ⇓ = Average or mean

1. Z - test
2. t - test
3. CHI Square ⇒ categorical data
4. Annova ⇒ Variance

We do statistical Analysis on Inferential Stats. Here we draw conclusion for a population after doing some experiments on "sample"

- Good Interview Question ⇒ when to use T-test vs Z-test.

Do you know the population std σ ?

- Yes
- No → Use T-test

Is sample size above 30 = ?

- Yes → Use Z-test
- No → Use T-Test

- To understand Z-test lets solve a problem

Q. The average height of all residents in a city is 168 cm with a population std σ = 3.9.

A doctor believes that the mean is different. He measured the height of 36 individuals and found

the height avg to be 169.5 cm

a.) state Null and Alternate hypothesis

b.) At 95% CI, is there enough evidence to Reject null hypothesis.

Ans :⇒ $\mu$ = 168 $\sigma$ = 3.9 $n$ = 36 $\bar{x}$ = 169.5

$\mu$ ↓ population mean

Sample size ↓ $n$ = 36

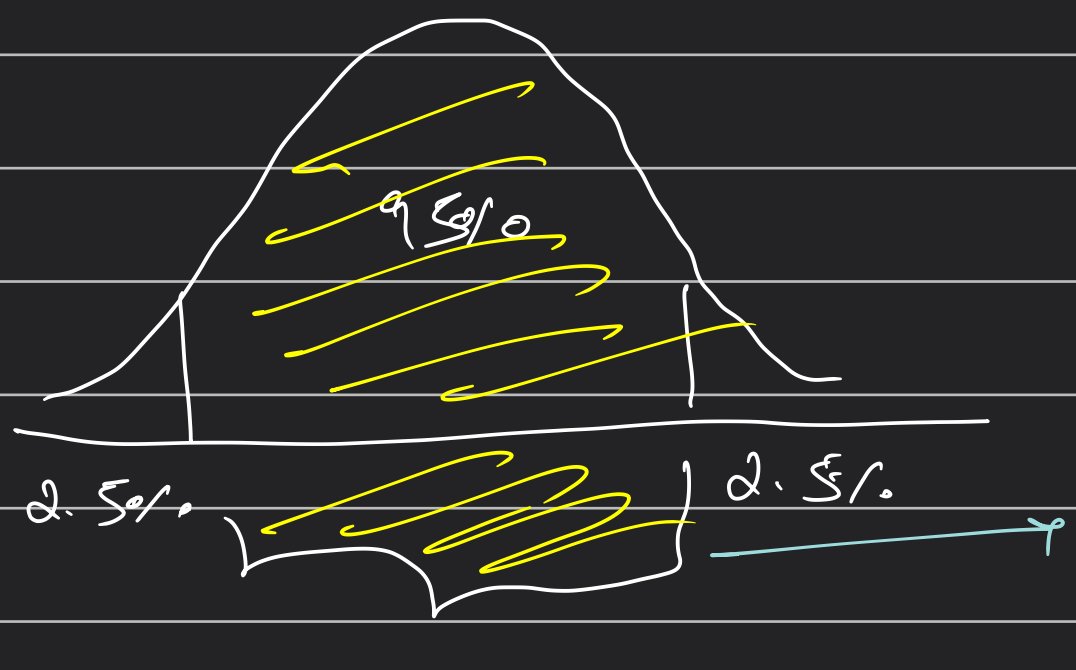
sample mean ↑ $\bar{x}$ = 169.5

a) null hypotheses $H_0$ $\mu = 168$ cm

b) Alternate hypothesis $H_1$ $\mu \neq 168$ cm } 2 Tail Test }

c) Confidence Interval = 0.95 ⇒ 95%

SI $= \alpha = 1 - 0.95 = 0.05$



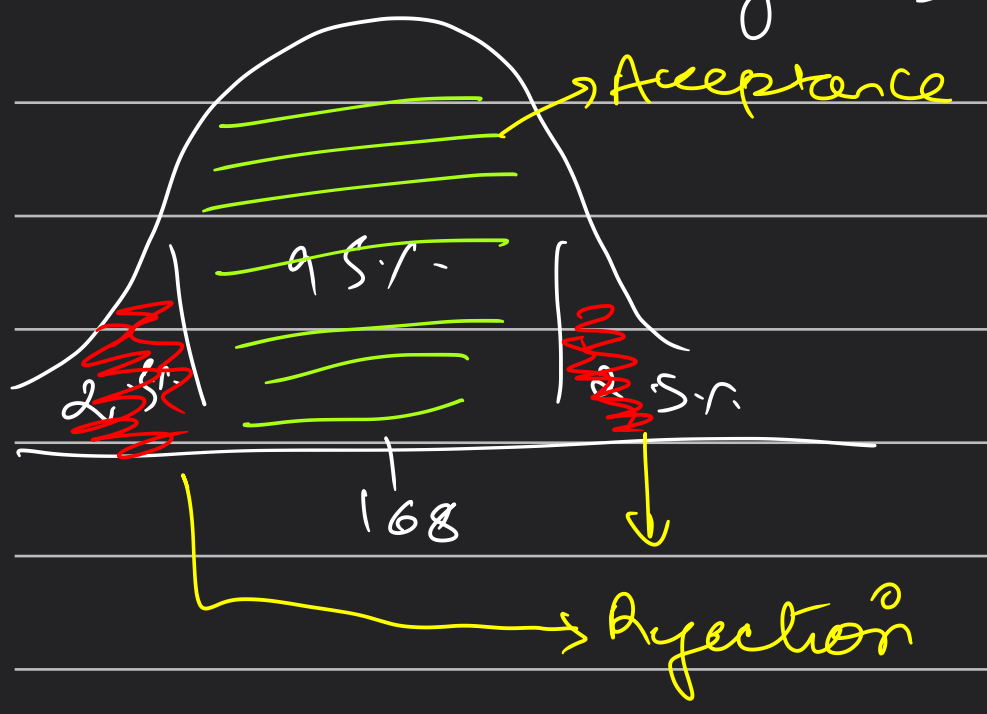

If our Z Test value comes in this 95% interval then we accept the Null hypothesis

or we fail to Reject the null hypothesis

But why 2-Tail Test ? Because the doctor believes the mean to different

to either mean is greater or lesser than 168 cm.

Decision Boundary ⇒




Now through Z-Test we need to find out where does this fall after statistical Analysis.

⇓

For this we need to find how much area under the curve does confidence interval covers. And we find that through Z-table

Let start from the right?

Full area = 1

Rejection on Right = 2.5%

So we need to find a Z table entry of

1 − 0.025 = 0.9750

So we will add the Z value & point value from Z table?

Z = 1.6 point = 0.06

Area Under Curve for 0.9750 = **+1.96**

We either check Right side (+ve) or left side (−ve) of Z table.

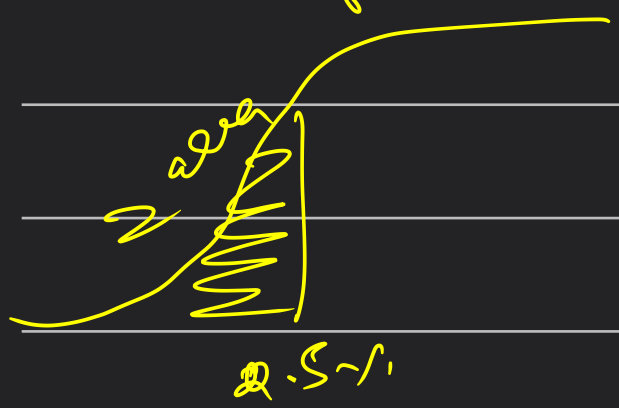
So particular the left side = **−1.96**

So what we are trying to say?

So If are Z test value occur between −1.96 and +1.96 then only we will

accept the Null hypothesis. otherwise we will reject it.

For left side calculation we can do




$= 1. - 0.9750 = 0.0250$

Now we do Z-Test

Formula for Z test

$$\text{Z. Test} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{169.5 - 168}{3.9/\sqrt{36}}$$

Standard Error? ← $\sigma/\sqrt{n}$

Z = 2.31

$$\text{ZScore} = \frac{\bar{x} - \mu}{\sigma}$$

↓

For Each Sample size

$n = 1$

conclusion ⇒ since $2.31 > 1.96$ we reject the Null hypothesis.

This means what?

⇒ This means that the Doctor was right that population mean was not right.

## Understanding T-Test {One sample T-Test}

① On the population the average IQ is 100. A team of researchers want to test a new medicator to see if it has either a +ve or -ve effect on intelligence or no effect at all.

A sample of 30 participants who have taken the medicator has a mean of 140 with a S.D of 20. Did medicine affect the intelligence?

CI = 95%

$\mu = 100 \quad n = 30 \quad \bar{x} = 140 \quad s = 20$

↓

Sample S.D

CI = 95% $\quad$ SV = 5% $= 0.05$

Null Hypothesis $H_0 : \mu = 100$

Alternate Hypothesis $H_1 : \mu \neq 100$ {2 Tail Test}

Significance value = 0.05

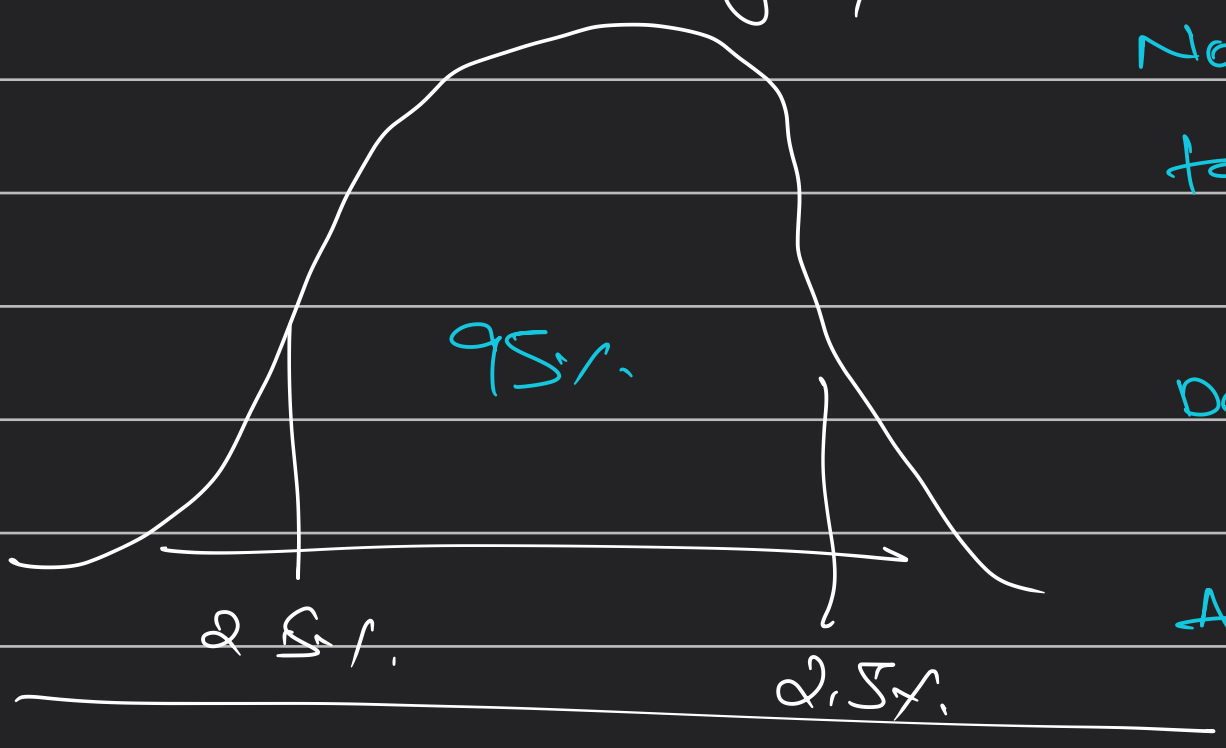
Do to T-Test we need one very imp thing at tests called "Degree of freedom"

Degree of freedom = $n - 1 = 29$

↓ $\quad$ ↓

Kind of choices Available $\quad$ Sample size

Decision Boundary



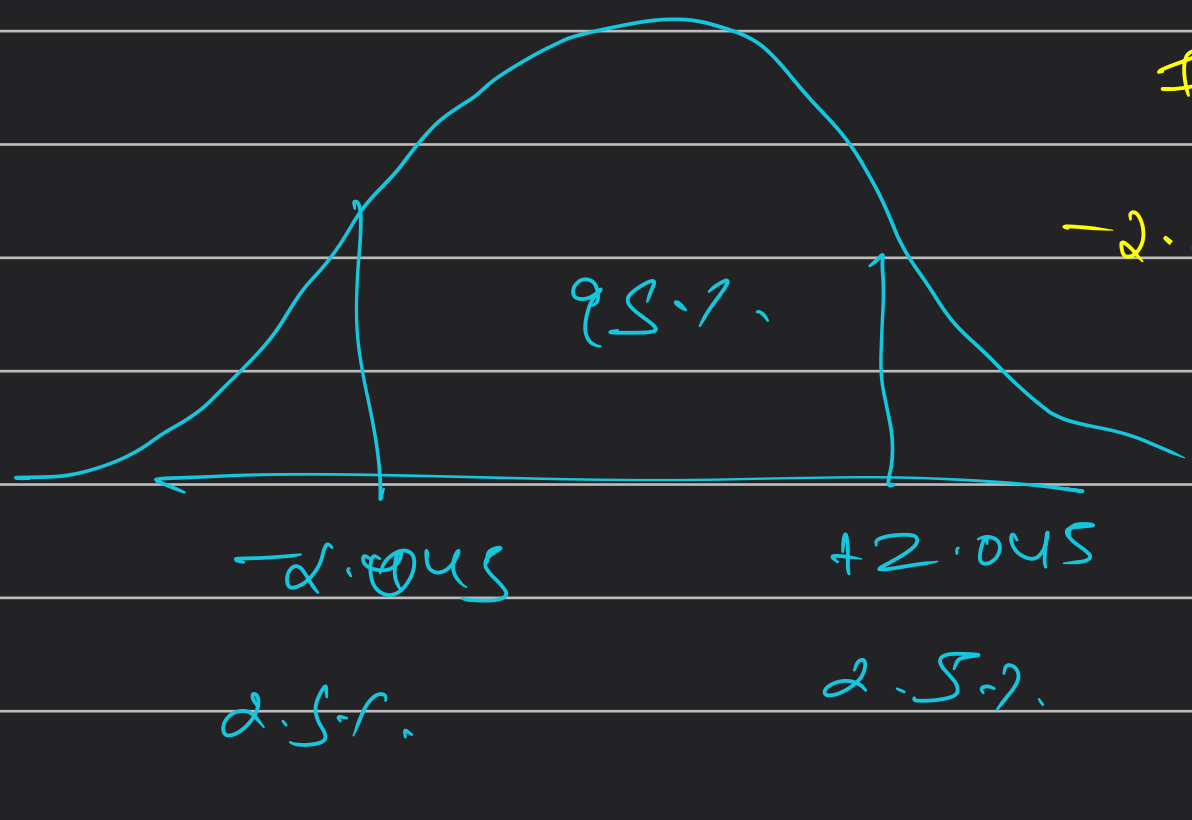

Now we go to t table for calculation

Degree of Freedom = 29

And Based upon One tail or two tail

As Two Tail we have 0.025

- Decision Boundary = 2.045




If value of T Test is -2.045 & greater than 2.045. Then we reject Null Hypothesis

Calculate the T-test statistics:

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}} = \frac{140 - 100}{20/\sqrt{30}} = 10.96$$

S.D sample

of conclusion $\Rightarrow$ 10.96 > 2.045 we reject the Null hypothesis. Now this T test is more no +ve side this means medicine has

+ve effect on IQ and if it has been on -ve side means medicine had -ve effects on IQ.

$t_{0.96}$ ⇒ t-test score ⇒ IQ increased.