DAY-5
Statistics DATA SCIENCE

Inferential Statistics

Making Assumption & conclusion when we have sample Data regarding Population Data

Agenda

1. Hypothesis testing
2. P-value
3. Confidence interval
4. Significance value

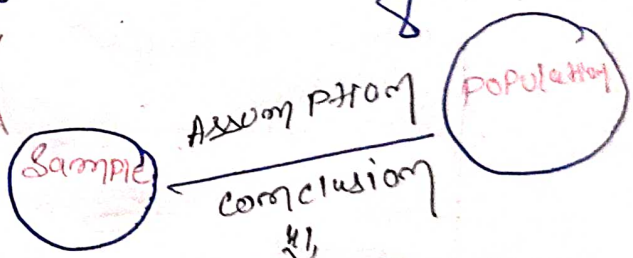
⇒ Hypothesis testing

In order to validate assumption & conclusion we use Hypothesis testing

types of Hypothesis testing

⇒ NULL Hypothesis (H_0)

⇒ Null Hypothesis (H_0) is default statement

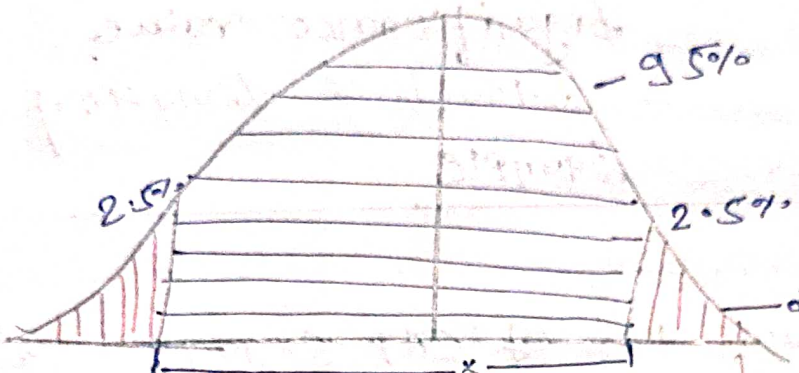


Alternative Hypothesis (H_1)

⇒ A Person is Criminal or not

1. Person is not Criminal
2. Person is Criminal
3. Experiment / Proof :- DNA / Finger Print / Weapon / eye witness

* Confidence Interval



Significance value :-

C.I = 95%.

Significance value (SV) = 1 − 0.95

= 0.05

* Point estimates :-

:- The value of any statistics that estimates the value of a parameter is called Point Estimate.

Statistic	Parameter
$\bar{x}$	μ — Population Mean
↓	
Sample mean	

Point Estimator ± Margin of Error = Parameter => Population Mean

Lower_CI = Point Estimate − Margin of Error

Higher_CI = Point Estimate + Margin of Error

Margin of Error $\Rightarrow Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

where,

α = significance value

σ = standard deviation

$\sqrt{n}$ = sample,

Question

1) On the Quant test of CAT Exam, a sample of 25 test taker has mean of 520 with population standard deviation of 100. So, construct a 95% CI about the mean.

$\Rightarrow n = 25 \quad \bar{x} = 520 \quad \sigma = 100$

$CI = 95\% \quad \alpha = 1 - 0.95 = 0.05$

Lower_CI = Point Estimater − Margin of Error

$= 520 - Z_{\frac{0.05}{2}} \left[\frac{100}{\sqrt{25}}\right]$

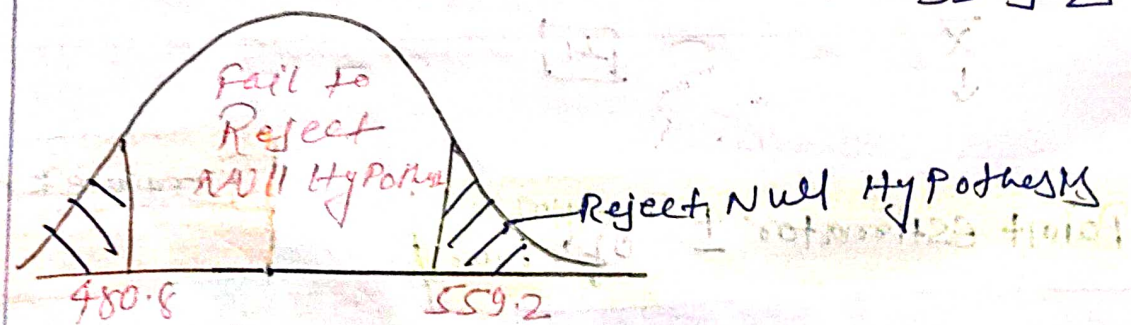
(value of Z is from Z table)

$= 520 - 1.96 \times 20$

$= 480.8$

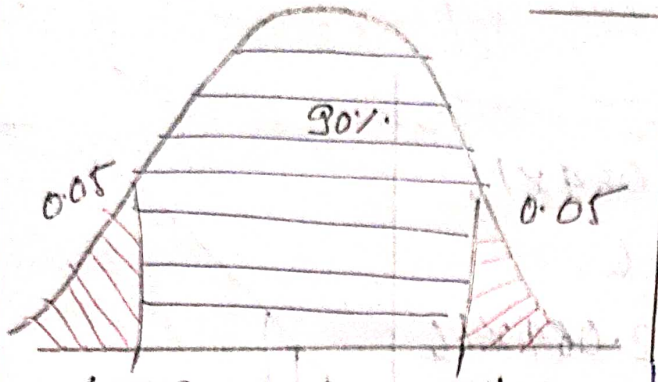
Higher_CI = Point Estimater + Margin of Error

$= 520 + 1.96 \times 20 \Rightarrow 559.2$



Q2. $\bar{x} = 480 \quad \sigma = 85 \quad n = 25 \quad C.I = 90$

$Sv(\alpha) = 1 - 0.90 = 0.10$



$$\text{Lower\_CI} = 480 - Z\frac{0.10}{2}\left[\frac{85}{25}\right]$$

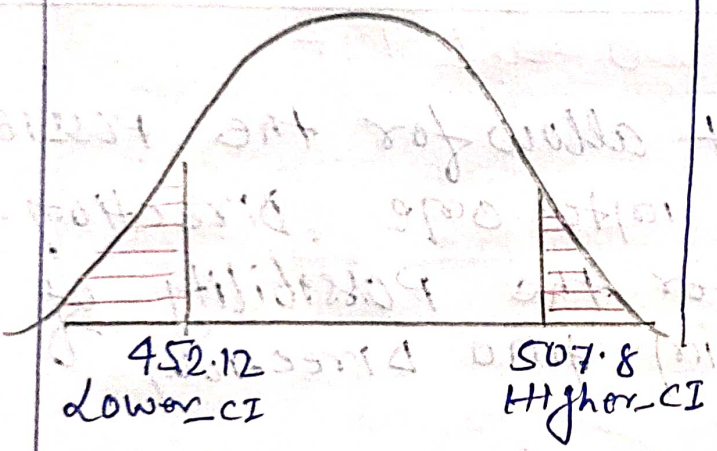
$$\Rightarrow 480 - Z\,0.05 \times 17$$

$$\Rightarrow 480 - 1.64 \times 17$$

$$\Rightarrow 452.12$$

$$\text{Higher\_CI} = 480 + 1.64 \times 17$$

$$\Rightarrow 507.8$$



Q3 On the Quant test of CAT Exam, a Sample of 25 test taker has mean of 520 with sample standard deviation of the Construct 95% CI about the mean?

$$\Rightarrow \bar{x} \pm t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$ → Because Sample Standard deviation is given

* For t-test we need to find degree of Freedom $= n - 1 = 25 - 1$

$= 24$

$$\text{Lower\_CI} = 520 - t_{0.5} \left(\frac{80}{\sqrt{25}}\right)$$

$$\text{Lower\_CI} = 520 - \frac{t_{0.5}}{2}$$

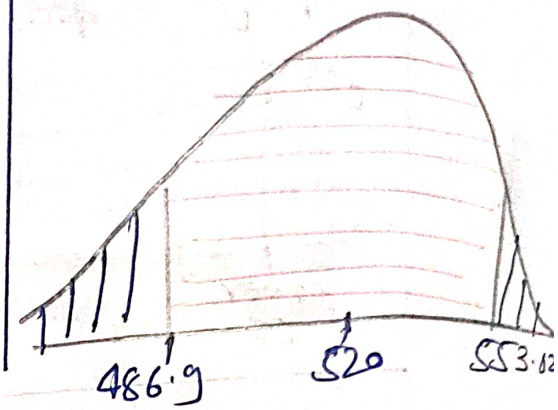
* We are taking value of t from t-table

$$= 520 - 2.064 \times 16$$

$$= 486.976$$

$$\text{Higher\_CI} = 520 + 2.064 \times 16$$

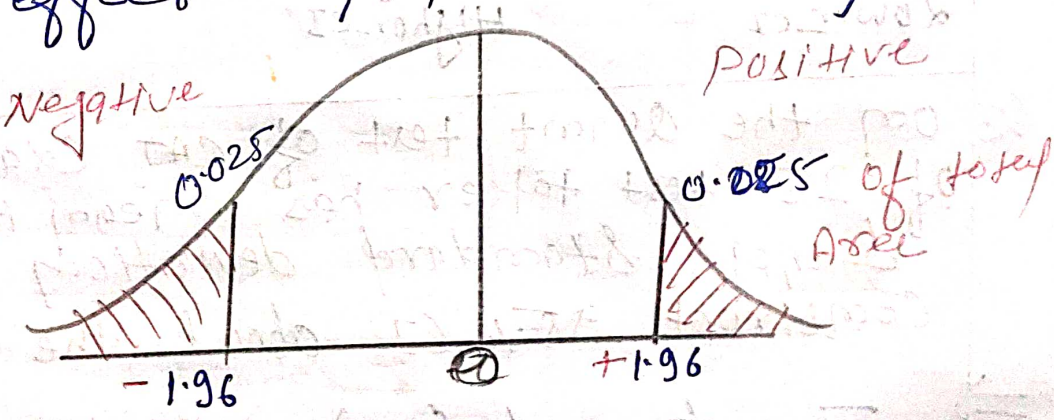
$$= 553.024$$



One tail & two tail test

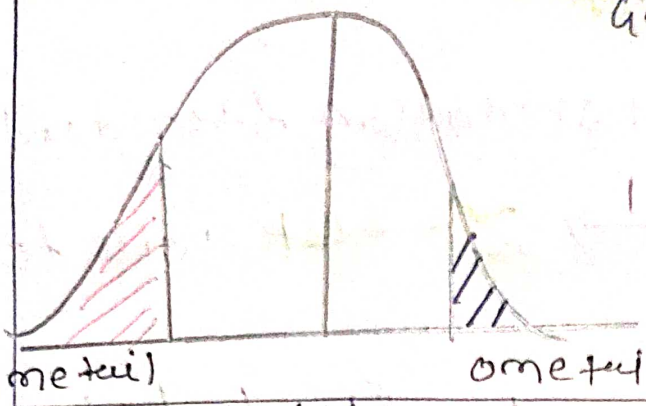
⇒ One-tail test allow for the possibility of an effect into one direction.

⇒ Two tailed for the possibility of an effect in two directions



Question

A college in two has 85% placement rate A new college was Recently open & it was found that Sample of 150 Student had a placement Rate of 88% with σ = 4
Does this college has a Different Placement rate with 95% C.I ?



Greater than 65% on Right tail

lesser than 65% on left tail

Assignment ↑

Hypothesis testing Problem

Q. A factory has a Machine that fills 80 ML to baby Medicine in a bottle. An employee believes that Average of baby Medicine is not 80 ML. using 40 Sample he Measured the 78 ML with Standard deviation of 2.5?

① State NULL & Alternative Hypothesis

② At 95% CI, is there enough evidence to support Machine is working Properly or not?

⇒ Step 1:

NULL Hypothesis $H_0 = \mu = 80$

Alternative Hypothesis $H_1 = \mu \neq 80$

Step 2:

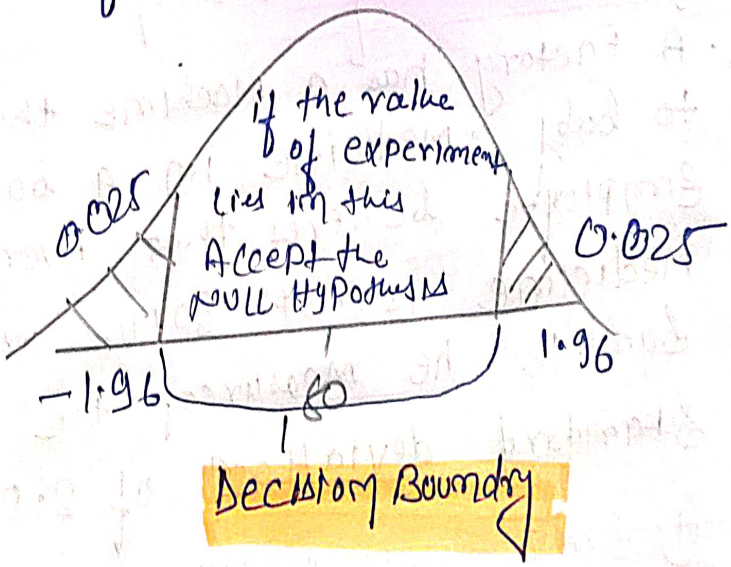
$\mu = 80$ $n = 40$ $\bar{x} = 78$ $S = 2.5$

C.I = 95 SV(α) = 0.05

Important Point

* $n \geq 30$ or Population standard deviation is given :→ **Z-test** will be applied

* $n < 30$ **AND** Sample standard deviation is given :- **t-test** will be applied



Calculate Z-test

$$Z = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

→ If σ given use σ — Standard error

$$Z = \frac{78 - 80}{25/\sqrt{40}} \qquad Z = -5.05$$

Conclusion

Decision Rule

⇒ if $Z = -5.05$ is less than -1.96 or greater than $+1.96$, Reject the NULL Hypothesis

Reject the H_0 { There is some Problem in Machine }

Q A Complain Registered, the boys in a Government School are underfeed. Average weight of Age 10 is 32kg with S.D = 9 kg. A Sample of 25 Boys were selected from Govt. school & Avg. weight was found 29.5 kgs. with CI = 95%. Check it is true or false

Step-1

⇒ NULL Hypothesis = $\mu = 32$

Alternative = $\mu \neq 32$

Step-2

$$Z\text{-test} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{29.5 - 32}{9/\sqrt{25}} = -1.39$$

Conclusion

$\therefore$ $-1.39 > -1.96$

Accept the Null Hypothesis at 95% CI

we fail to Reject Null Hypothesis

* The Boys feel well