# Inferential Statistics

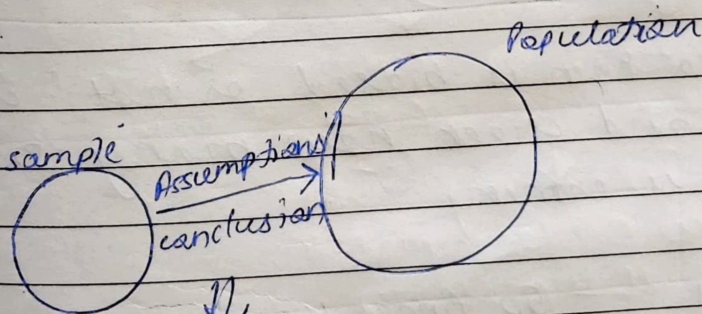
→

1) Hypothesis testing
2) P-value
3) confidence Interval
4) significance value

→ Z test
→ t test
chi square test
Annova test (F-test)

Bernoullis
Binomial } Transformation
Powers law

* Inferential statistics




⇓

Hypothesis testing

Based on sample data, we are making assumptions/ conclusions of population data.
And these assumptions is verified through hypothesis testing.

Page No.:
Date:

# Steps of Hypothesis Testing

① Null Hypothesis :-

→ we will use default value.

Example :-
If we have to decide whether coin is fair or not, we will initially assume that coin is fair.

② Alternate Hypothesis :-

→ Example :-

coin → fair / ~~fa~~ unfair (check).
Initial Assumption will be coin is not fair.

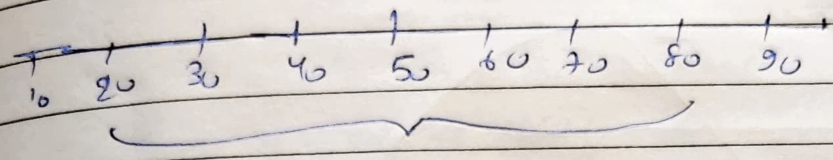
③ Perform Experiment :

checking a coin is biased or not. we toss it 100 times and check how many sides a Head / Tail is coming.

100 times — 50 times Head
(total) 60 times Head
70 times Head

↓

we will then consult domain expert about this situation and then a range will be decided by him called as Confidence Interval (C.I) and based on this scale we can make to conclusion




CI ⇒ confidence interval

⇒ If we fail to reject the Null Hypothesis [Inside CI]
⇒ we reject the Null hypothesis [Outside CI]

⇒ Example: -2
Person is criminal or not {Murder case.}

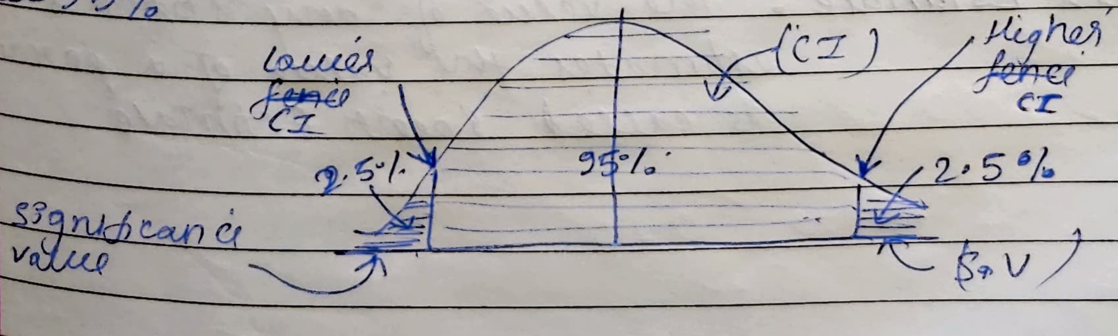
1) Null hypothesis :- Person is not criminal
2) Alternate hypothesis - Person is criminal
3) Experiment / Proof :- DNA, finger print, weapons, eye witness etc.

⇓

Judge

# Confidence Interval (CI)

CI = 95%



# Significance value

$\boxed{S.V. = 1 - \text{Confidence Interval}}$

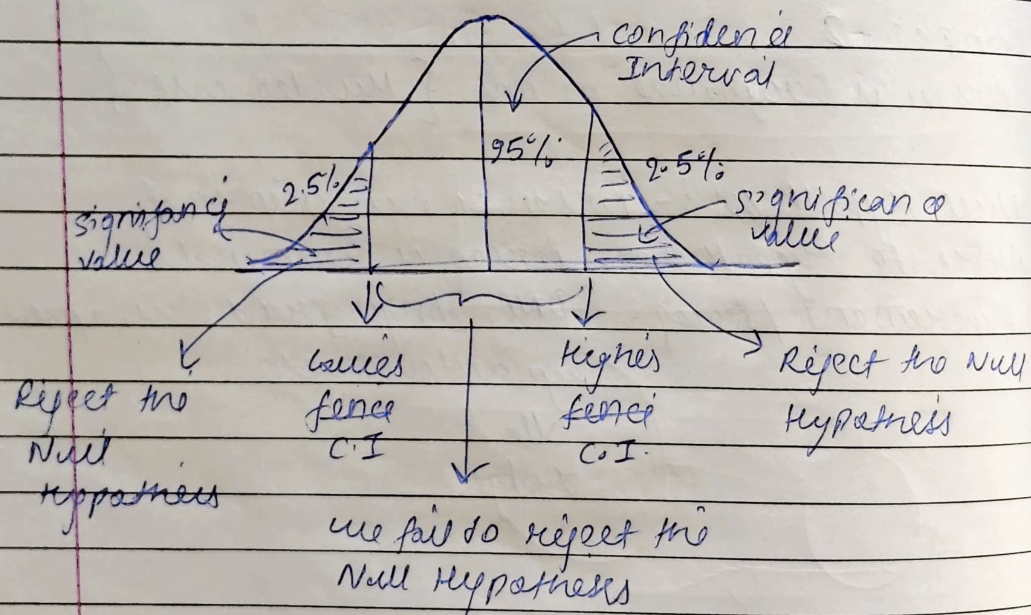
$1 - 0.95$

$= 0.05$

~~CV~~

⇒ Confidence Interval and Significance value is decided by Domain Expert

Normally CI will be small




V.Imp.

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

Page No.:
Date:

sample mean $\bar{x} \longrightarrow \mu$ Population mean

statistics Parameters

$\bar{x} \geq \mu$ or $\bar{x} \leq \mu$, so

$$\boxed{\text{Point Estimate} \pm \text{Margin of Error} = \text{Parameter}}$$

Point Estimate → sample mean

Parameter → Population mean

$$\boxed{\begin{array}{l} \text{Lower C.I} = \text{Point Estimate} - \text{Margin of Error} \\ \text{Higher C.I} = \text{Point Estimate} + \text{Margin of Error} \end{array}}$$

# Margin of Error = $Z_{\alpha/2} \boxed{\frac{\sigma}{\sqrt{n}}}$ ⇒ standard error

$$\boxed{Z_{\alpha/2} \frac{\sigma}{\sqrt{2}}}$$

$\alpha \rightarrow$ Significant value

Q) On the quant test of CAT Exam, a sample of 25 test takers has a mean of 520 with a ~~sample~~ population standard deviation of 100. Construct a 95% CI ~~abou~~ about the mean?

Soln:

$n = 25$

$\bar{x} = 520$

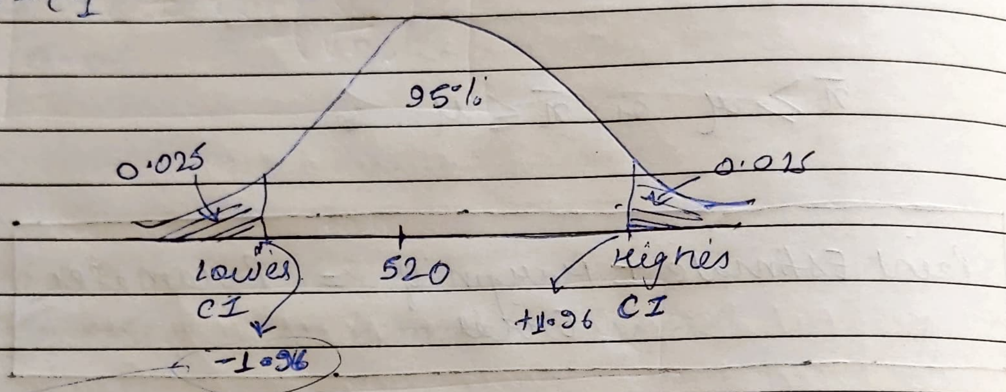
$\sigma = 100$

$CI = 95\%$

$S.V = 1 - C.I$

$= \cancel{1 - 0.9} \quad 1 - 0.95 = 0.05$

$\alpha = 1 - CI$




Lower CI = Point Estimate − Margin of Error

$= 520 - Z_{\frac{0.05}{2}} \frac{\sigma}{\sqrt{n}}$

$= 520 - Z_{0.025} \frac{\overset{20}{\cancel{100}}}{\cancel{5}} = 520 - 1.96 \times 20$

$= 480.8$

we find this value by referring to Z table and we will have to check that when area under curve comes 0.025

Higher CI $= 520 + 1.96 * 20$

$= 559.2$

(2) $\bar{x} = 480$

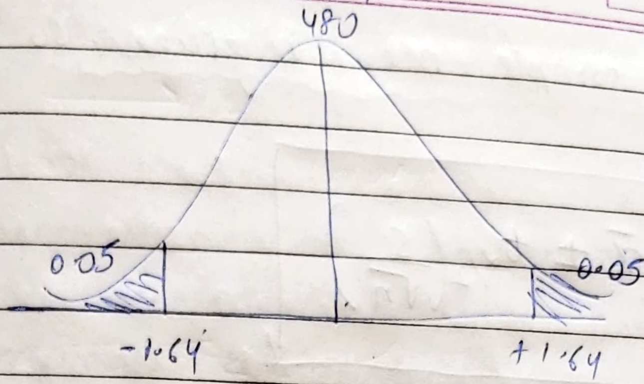
$\sigma = 85$

$n = 25$

$CI = 0.90$ or 90%

Soln $S.V = 1 - 0.90 = 0.10$

$\alpha = 0.10$




Lower CI = Point Estimate − Margin of Error

$= 480 - Z_{\frac{0.10}{2}} \frac{\sigma}{\sqrt{n}}$

$= 480 - Z_{0.05} \frac{85}{\sqrt{25}}$

$= 480 - 1.64 \times \frac{\cancel{85}^{17}}{\cancel{5}}$

$= 480 - 27.88$

$= 452.12$ (Ans)

Higher CI $= 480 + 27.88$

$= 507.88$ (Ans)

(Q) On the Quant test of CAT exam, a sample of 25 test takers has a mean of 520, with a sample standard deviation of 80. Construct a 95% CI about the mean?

Ans

$\bar{x} = 520$

$s = 80$

$CI = 95\%$

$SV = 1 - 0.95 = 0.05$

★ ⇒ Here sample mean is given, so

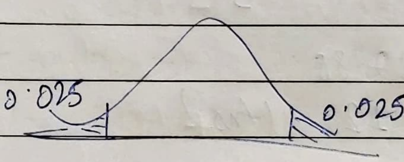
$$\boxed{\bar{x} \pm t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)}$$

Here we use t test

$$\boxed{\text{Degree of freedom} = n - 1}$$

how many choices do we have

$25 - 1 = 24$




Refer to t table - check the degree of freedom (D

$$\text{Lower CI} = 520 - t_{0.05/2}\left(\frac{80}{5}\right)$$
$$= 520 - t_{0.025}(16)$$
$$= 520 - 2.064 * 16$$
$$= 486.976$$

$$\text{Higher CI} = 520 + 2.064 * 16$$
$$553.024$$

① 1 tail and 2 tail test

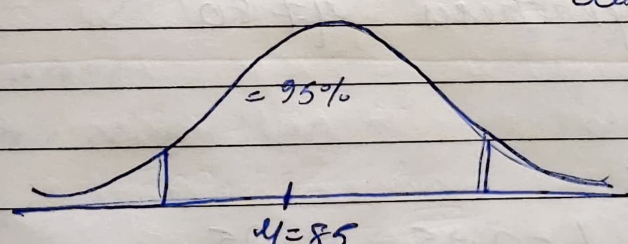
* college in Town A has 85% placement rate. A new college was recently opened and it was found that a sample of 150 students had a placement rate of 88% with a standard deviation of 4%. Does this college has a different placement rate with 95 C.I ?

⇓

two tail

placement rate ~~greater~~ greater than 85% } 1 tail
" " less " "

Different placement rate (either greater or lesser) - so we will focus on both tail
∴ 2 tail test.




# NOTE -

→ Z test - Population standard deviation is known or ($n > 30$) and data belongs to normal distribution

→ T test - Population standard deviation is unknown
- distribution - normal
- sample size $< 30$