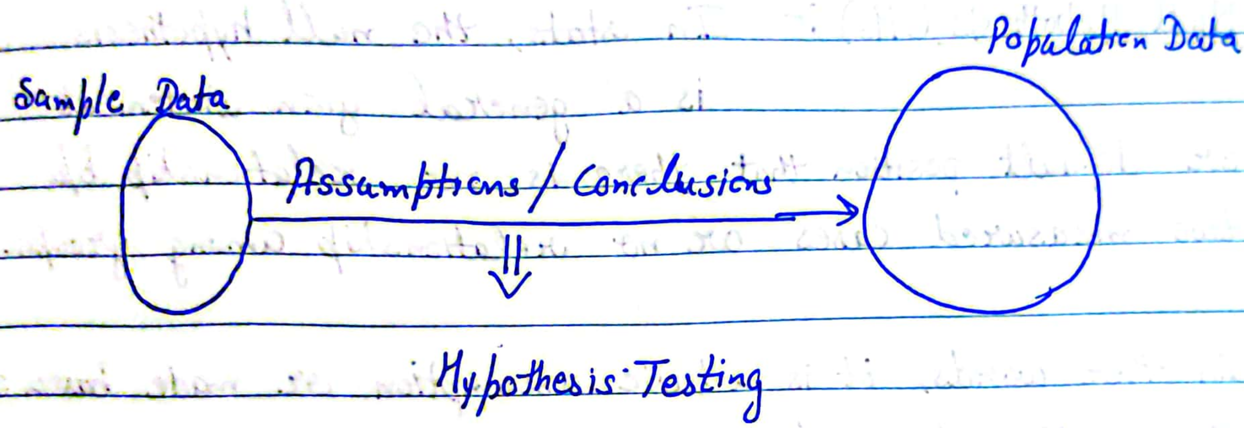
Day - 5

Inferential Statistics

→ Based on sample data we make assumptions or ~~on~~ conclusions about the population data.

→ We use Hypothesis Testing to validate these conclusions.



Hypothesis Testing

→ Hypothesis are statements about the given problem.

→ Hypothesis Testing is a statistical method that is used in making a statistical decision using experimental data.

→ Hypothesis Testing is basically an assumption that we make about a population parameter. It evaluates two mutually exclusive statements about a population to determine which statement is best supported by the sample data.

→ Why do we use Hypothesis Testing?

A Hypothesis Test evaluates ~~how~~ two mutually exclusive statements about a population to determine which statement is best supported by the data.

→ **Steps of Hypothesis Testing :-**

① Null Hypothesis (H_0) :- In stats, the null hypothesis is a general given statement or default position that there is no relationship b/w two measured cases or no relationship among groups.

→ In other words, it is a basic assumption or made based on the problem knowledge.

→ By default, Null Hypothesis will always be True.

⇓

[Experiment: if a coin is fair or not]

⇓

$P(H) = 0.5$; $P(T) = 0.5$

⇓

Null Hypothesis :- Coin is fair (by default)

② Alternate Hypothesis :- Opposite of Null Hypothesis

⇓

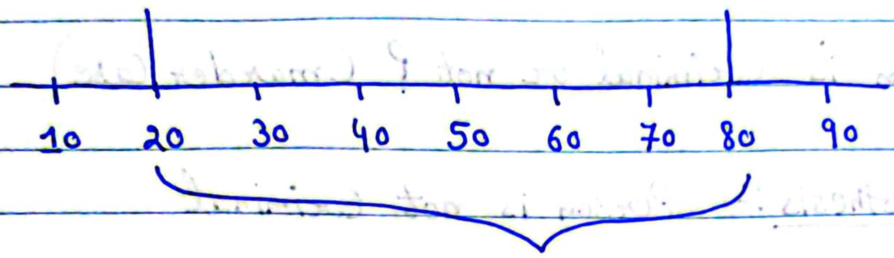
Coin is not fair

③ Perform Experiments :-

E.g. :- We Tossed coin 100 Times

⇓

Let say $\mu = 50$, $\sigma = 10$



C.I. → Confidence Interval

→ 100 Times ⇒ 50 Times Head

or

60 Times ~~Head~~ Head

50 - 50		
60 - 40	}	Fair coin
70 - 30		

As my C.I range is 20-80 → 80:20 is unfair

⇓

If Head comes less than 20 times or greater than 80 times consider it unfair.

→ Let say Head comes 75 times [C.I. = 20-80]

↓

Null Hypothesis is True and ~~&~~ we accept Null Hypothesis

↓

We fail to reject Null Hypothesis.

If outcome is outside C.I

⇓

We reject the Null Hypothesis.

→ Fail to reject Null Hypothesis (inside CI Range)
Reject Null Hypothesis (outside CI Range)

} Conclusion

E.g. → 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, Fingerprints, CCTV etc

⇓

Conclusions

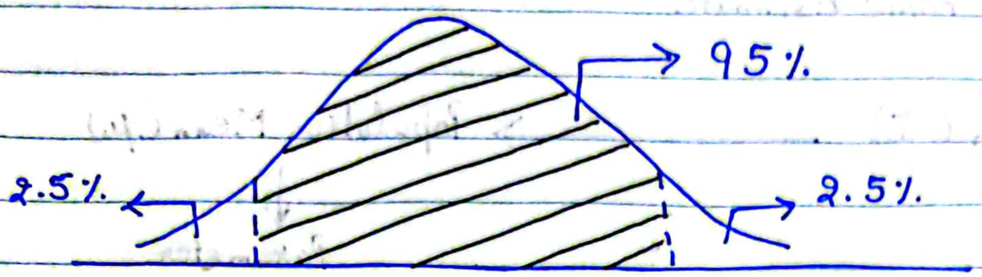
Confidence Interval

A C.I., in statistics refers to the probability that a population parameter will fall b/w two set values for a certain proportion of times.

C.I measures the degree of uncertainity or certainity in a ~~sample~~ sampling method. A confidence ~~uncertainity~~ interval can take any number of probabilities, with the most common ~~bein~~ being a 95% or 99% confidence Interval.

→ **95% Confidence Interval**

⇓



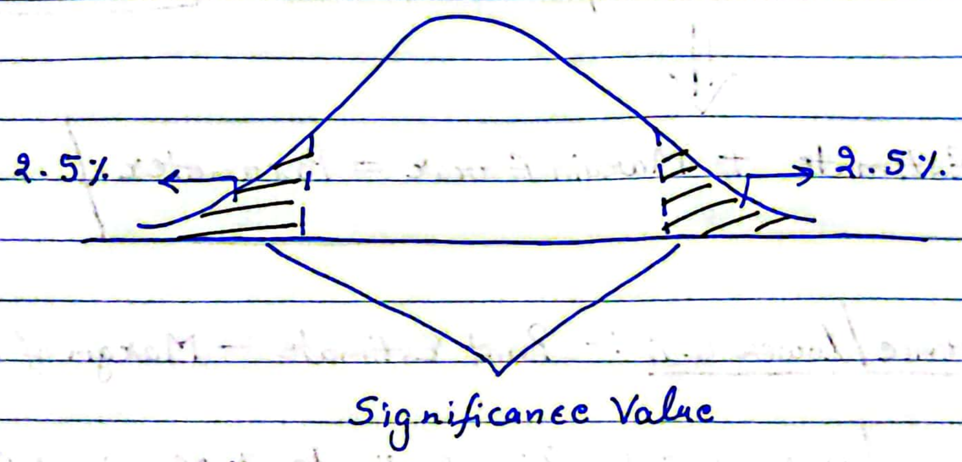
→ If C.I. = 95%

⇓

~~Significance Level (S.~~

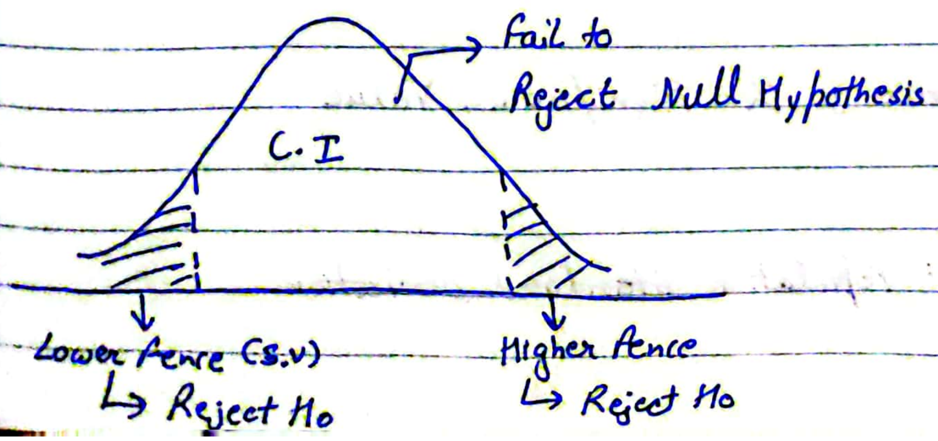
Significance Value (SV) → $1 - 0.95$

$= 0.05$



⇒ C.I and S.V are defined by Domain Experts.

E.g:- For vaccinne trial or medical problems ; C.I. will be very less.



$\boxed{S.V = 1 - C.I}$

(3) **Point Estimate** :- The value of any statistics that estimates the value of a parameter is called Point Estimate

Sample mean ($\bar{x}$) $\longrightarrow$ Population Mean (μ)

$\downarrow$ Statistic $\downarrow$ Parameter

With the help of sample mean we are estimating the value of population mean.

$\rightarrow$ Parameter $\Rightarrow$ Population mean (μ)

Point Estimate $\Rightarrow$ Sample mean ($\bar{x}$)

$$\boxed{\begin{array}{c}\bar{x} \geq \mu \\ \bar{x} \leq \mu\end{array}}$$

$\Downarrow$

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

$\rightarrow$ Lower fence / Lower C.I :- Point Estimate $-$ Margin of Error

Higher fence / Higher C.I :- Point Estimate $+$ Margin of Error

$$\boxed{\text{Margin of Error} = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}}$$

$\frac{\sigma}{\sqrt{n}}$ = Standard Error ; α = Significance Value

σ = Population Standard Deviation.

Example :- On the quant test of CAT Exam, a sample of 25 test takers has a mean of 520 with a population standard deviation of 100.

Construct a 95% C.I about the mean?

Sol :-

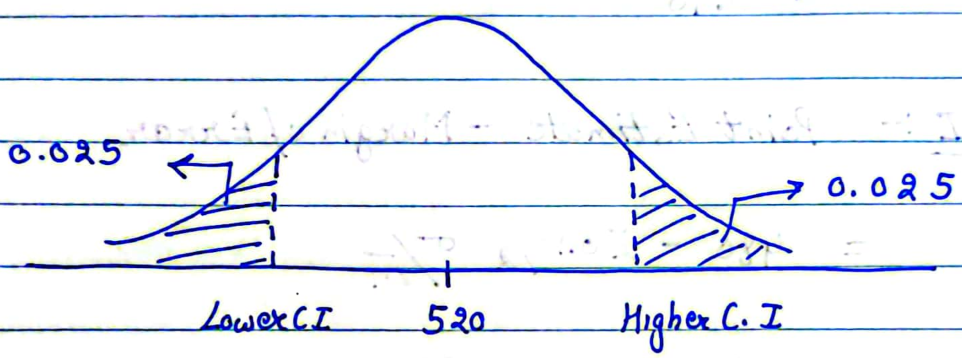
$n = 25$

$\bar{x} = 520$

$\sigma = 100$

$C.I = 95\% = 0.95$

$\Rightarrow$ $S.V = 1 - 0.95 = 0.05$



$\Rightarrow$ Lower C.I. = Point Estimate − Margin of Error

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

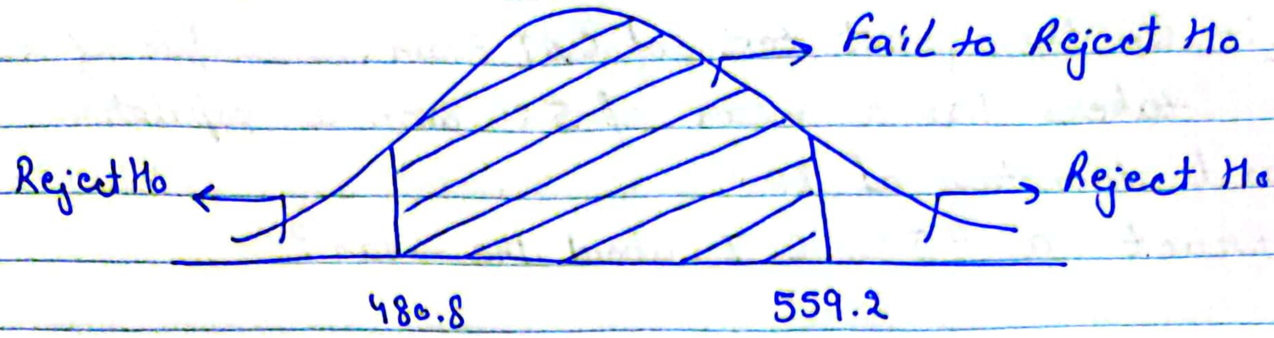
$= 520 - Z_{0.025} \dfrac{100}{5}$ ∴ Z-Calculation from Z-Table

$= 520 - 1.96 \times 20$

$= 480.8$

Higher C.I : $= 520 + 1.96 \times 20$

$= 559.22$



E.g. 2 :- $\bar{x} = 480$, $\sigma = 85$, $n = 25$, C.I = 90%.

Sol. $\Rightarrow$ S.V = $1 - 0.90$
$= 0.10$

$\rightarrow$ **Lower C.I** :- Point Estimate − Margin of Error

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

$= 480 - Z_{0.05} \frac{85}{5}$

$= 480 - 1.64 \times 17$
$= 452.12$

Higher C.I :- $480 + 1.64 \times 17$
$= 507.8$

$\Downarrow$

C.I. Range $\rightarrow$ [452.12 $\longleftrightarrow$ 507.8]

E.g. 3 :- 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 95% C.I about the mean?

Sol:- $\bar{x} = 520$, $s = 80$, C.I = 95% $\Rightarrow$ S.V = 0.05

If sample standard deviation is given

$\Downarrow$

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

t - test

$\downarrow$

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

$\Rightarrow$ Degree of freedom $= 25-1$
$= 24$

Lower C.I $= 520 - t_{0.05/2}\left(80/5\right)$

$= 520 - (2.064 \times 16)$

$= 486.976$

Higher C.I $= 520 + t_{0.05/2}\left(\frac{80}{5}\right)$

$= 520 + (2.064 \times 16)$

$= 553.024$

Hypothesis Testing Problems

Q A factory has a machine that fills 80 ml of Baby medicines in a bottle. An employee believes that the average amount of baby medicine is not 80 ml. Using 40 samples he measures the average amount dispersed by the machine to be 78 ml with a standard deviation of 2.5

(a) ~~Sta~~ State Null and Alternate Hypothesis
(b) At 95% CI, is there enough evidence to support if machine is working properly or not?

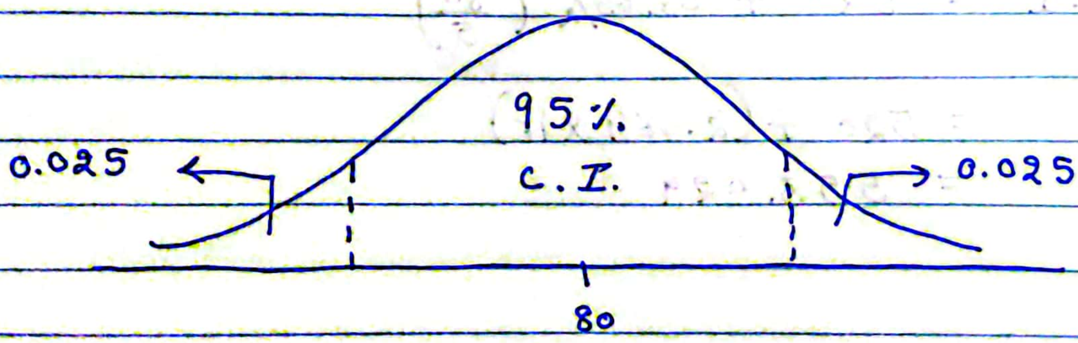
Sol :- Step 1 :- Null Hypothesis (H_0) $\longrightarrow$ $\mu = 80$
$\downarrow$
Machine working properly

Alternate Hypothesis (H_1) : $\mu \neq 80$ $\rightarrow$ Machine not working properly.

Step 2 :- $\mu = 80$, $n = 40$, $\bar{x} = 78$, $s = 2.5$

C.I = 95%

$\Rightarrow$ S.V (α) $= 1 - 0.95$
$= 0.05$



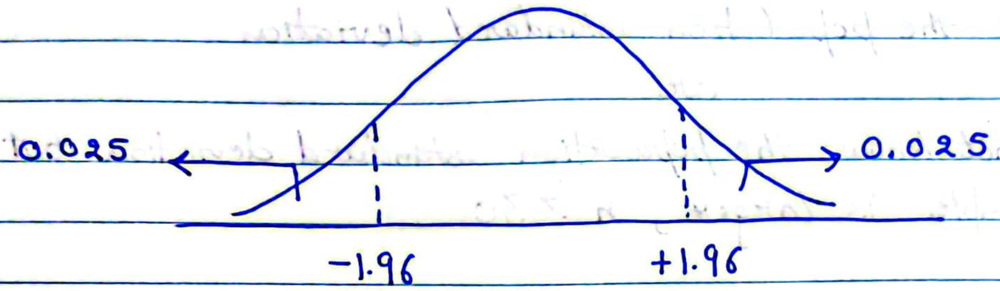
Step 3 :- $n = 40, \; s = 2.5$

when $n \geq 30$ or population std. dev $\rightarrow$ Z-Test
$n < 30$ and population std. dev $\rightarrow$ t-Test

$\Downarrow$

Here we will use Z-Test

Step 4 :- Perform the experiment



Step 5 :- Calculate Test statistics (Z-Test)

$$Z = \frac{\bar{x} - \mu}{\boxed{s/\sqrt{n}}} \rightarrow \text{Standard Error}$$

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

Step 6 :- Conclusions

Decision Rule :- If Z (-5.05) is less than -1.96 or greater than +1.96, Reject the Null Hypothesis. with C.I. of 95%.

$\downarrow$

Reject Null Hypothesis

$\downarrow$

There is some fault in the machine.

Q A complain was registered, the boys in a govt. school are underfed. Average weight of the boys of age 10 is 32 kgs with standard deviation 9 kgs. A sample of 25 boys were selected from the govt. school and the average weight was found to be 25.9 kgs. with C.I 95%. check if its True or False?

→ Conditions for Z-Test

① We know the population standard deviation

or

② We do not know the population standard deviation but our sample is larger, $n \geq 30$

→ Conditions for t-Test

① We donot know the population standard deviation.
② Our sample size is small, $n < 30$
③ Sample standard deviation is given.

Sol:- $\mu = 32$, $\sigma = 32$ kgs, $n = 25$

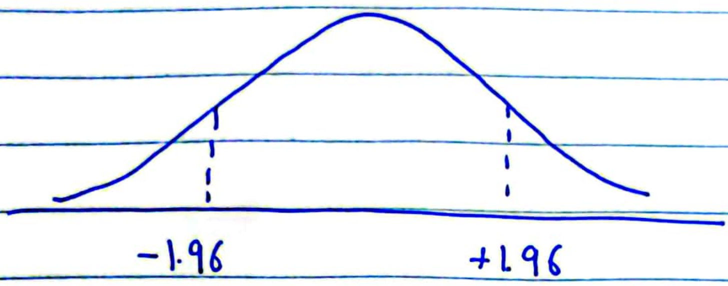
① $H_0 : \mu = 32 \rightarrow$ Fed Well

$H_1 : \mu \neq 32 \rightarrow$ Not Fed well

(2) C.I = 0.95

$\alpha = 1 - 0.95 = 0.05$

(3) Z-Test → 1 Tail Test



(4) $$\text{Z-score} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

$$= \frac{29.5 - 32}{9/\sqrt{25}}$$

$$= -1.39$$

(5) As $-1.39 > -1.96$

⇓

We accept Null Hypothesis with 95% C.I

⇓

We fail to Reject Null Hypothesis.