DAY 5 - Statistics.

Inferential Statistics:-

1. Hypothesis Testing.
2. P-value
3. Confidence Interval
4. Significance value

Tests :-

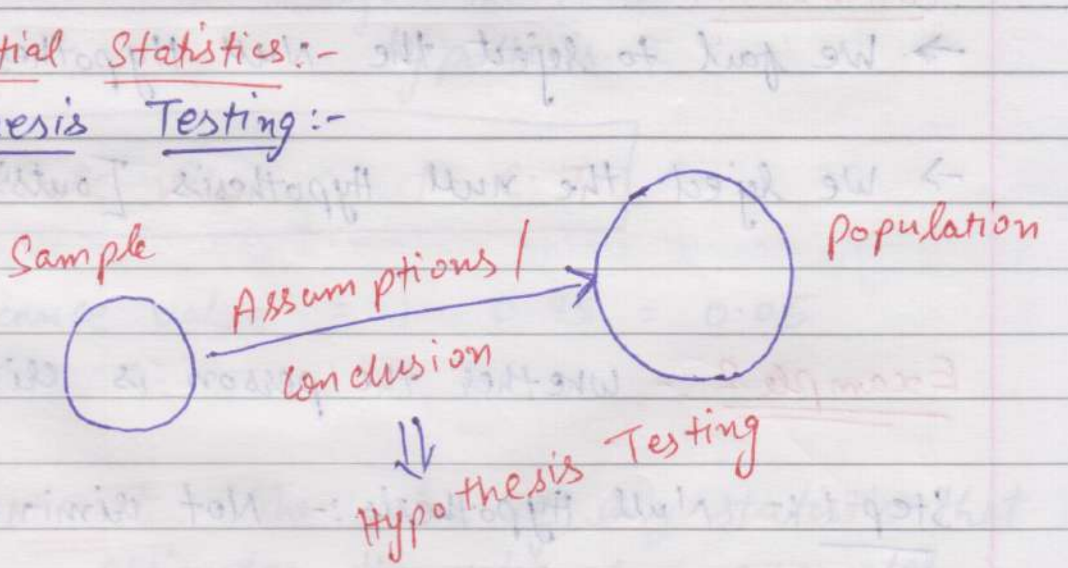
z test
t test
Chi square test
Annova test (F-test)

3 Distribution

1. Bernoullis
2. Binomial
3. Power Law

} TRANSFORMATION

★ Inferential Statistics:-
★ Hypothesis Testing:-



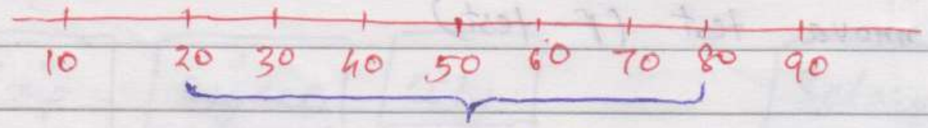
⇒ Steps of Hypothesis Testing:-

① Null Hypothesis :- Coin is fair

② Alternate Hypothesis :- Coin is not fair.

③ perform Experiments :-

Suppose we toss a coin and 50 times it shows head then its fair 50:50, head or tail ratio is fair. 60:40 or even 70:30, head and tail ratio is fair but 90:10 is not fair.



C.I → ~~tof~~ Confidence Interval.

A confidence Interval is the mean of your estimate plus and minus the variation in that estimate.

Conclusion:-

→ We fail to reject the Null Hypothesis [Within C.I]

→ We reject the null Hypothesis [outside C.I]

Example 2 :- whether the person is criminal or Not. (murder Case)

Step 1 :- Null Hypothesis :- Not criminal.

Step 2:- Alternate Hypothesis:- Person is criminal.

Step 3:- Perform Experiments / Proof :-

Collect DNA, Finger Prints, eye witness, footages.

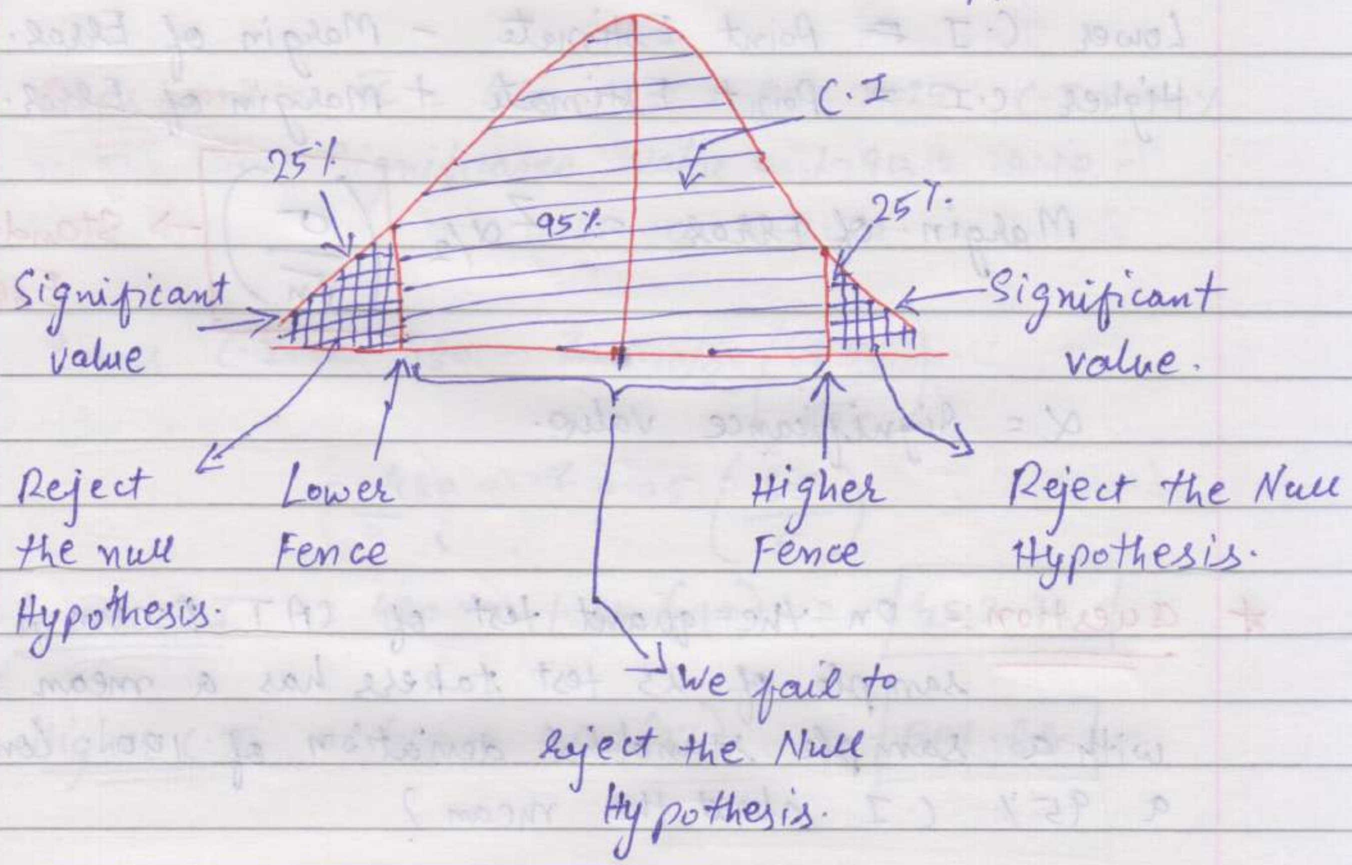
⇓

Judge will decide.

↳ Conclusion.

★ Confidence Interval [C·I]

Suppose C·I = 95%



$$\boxed{\text{Significante value} = 1 - C\cdot I}$$

Significance value $= 1 - 0.95 = 0.05$

Point Estimate :- The value of any statistic that estimates the value of a parameter is Called Point Estimate.

Sample mean ($\bar{x}$) $\xrightarrow{\text{Estimate}}$ Population mean (μ)

$\bar{x}$ - Statistic $\quad\quad$ $\mu \rightarrow$ Parameter

$$\boxed{\begin{array}{c}\text{Parameter}\\ \text{(population mean)}\end{array}} = \boxed{\text{Point Estimate}} \pm \boxed{\text{Margin of Error.}}$$

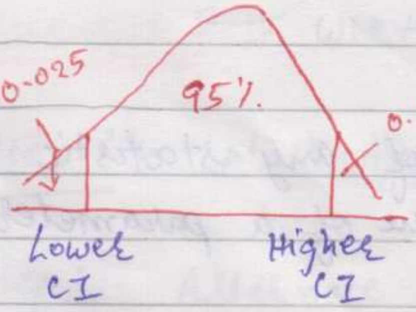
Lower C.I = Point Estimate - Margin of Error.
Higher C.I = Point Estimate + Margin of Error.

$$\text{Margin of Error} = z_{\alpha/2} \boxed{\left(\frac{\sigma}{\sqrt{n}}\right)} \rightarrow \text{Standard Error.}$$

α = Significance value.

★ Question:- On the quant test of CAT Exam, a sample of 25 test takers has a mean of 520 with a sample standard deviation of 100. Construct a 95% C.I about the mean?

Ans $\quad n = 25 \quad \bar{x} = 520 \quad \sigma = 100 \quad C.I = 95\%$.
Significance value = 1 - C.I = 0.05 = α



Lower CI = Point Estimate - Margin of Error.

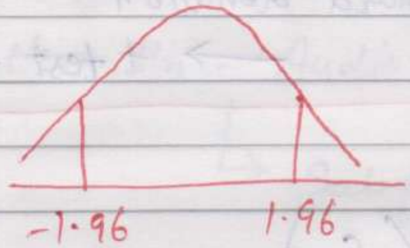
$$= 520 - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

$$= 520 - Z_{0.05/2}\left(\frac{100}{\sqrt{5}}\right)$$

$$= 520 - Z_{0.025}\left(\frac{\cancel{100}^{20}}{\cancel{5}}\right)$$

Lower C.I = 520 − $Z_{0.025}$ (20)

According to Z table 0.025 = −1.96.



$\therefore 520 - (1.96 \times 20) = \boxed{480.8}$

Higher C.I = 520 + (1.96 × 20)
= 559.2

* Question 2 :- $\bar{x} = 480$ $\sigma = 85$ $n = 25$ C.I = 90%.
Significance value = 1 − 90 = 0.10
$\alpha = 0.10$.

Lower C.I = $480 - Z_{0.10/2} \left(\frac{85}{\sqrt{25}}\right)$

$= 480 - Z_{0.05} \left(\frac{85}{5}\right)$

Lower C.I. = 480 − 1.64 (17) = $\boxed{452.12}$

Higher C.I = 480 + 1.64 (17) = $\boxed{507.88}$

* Question 3 :- On the quant test of CAT exam, sample of 25 test takers has a mean of 520, with a sample standard deviation of 80. Construct 95% C.I about the mean?

Ans :- $\bar{x} = 520$ S = 80 C.I = 95% S.V = 1 − 0.95 = 0.05
n = 25

~~CI~~

Note :-

→ If $n \geq 30$ or population standard deviation → Z test.

→ If $n < 30$ or sample standard deviation → t test.

$$C.I = \bar{x} \pm t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

(Confidence Interval)

To Find the value in t table we require Degree of freedom which is nothing but $n-1$

Degree of freedom $= n-1 = 25-1 = 24$

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

$$= 520 - t_{0.025}\left(\frac{80}{5}\right)$$

Now According to t table:

$$\boxed{t_{0.025} = 2.064}$$

Lower C.I $= 520 - (2.064 * 16) = 486.976$

Higher C.I $= 520 + (2.064 * 16) = 553.024$

★ 1 Tail and 2 Tail Test.

★ Questions :- Hypothesis Testing.

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

Ans. (a) State null

(b) Alternate Hypothesis.

(c) At 95% C.I, is there enough evidence to support machine is working properly or not.

Here :- $\mu = 80ml$ $n = 40$ $\bar{x} = 78ml$ $s = 2.5$

Null Hypothesis says $\mu = 80$
Alternate Hypothesis says $\mu \neq 80$ } Step 1

When Alternate Hypothesis does not agree with Null Hypothesis then we have to perform 2 Tail Test.

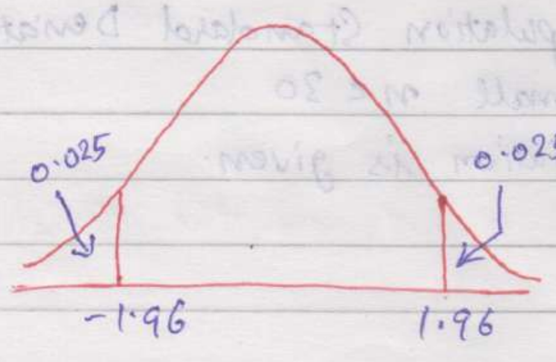
Step 2 :- $C.I = 0.95$ $S.V(\alpha) = 1 - 0.95 = \boxed{0.05}$

Step 3 :-

In this problem we will use Z test because $n = 40$ (means ≥ 30)
$s = 2.5$

$\alpha = 0.05$

$Z_{\alpha/2} = Z_{0.025}$
$= -1.96$



Calculate Z- test statistics

$$\boxed{Z\ score = \frac{\bar{x} - \mu}{\boxed{S/\sqrt{n}}}}$$

Standard Error.

$$Z = \frac{78 - 80}{\frac{2.5}{\sqrt{40}}} = \frac{(-2) \times 6.32}{2.5}$$

$$\boxed{Z = -5.05}$$

* Conclusion:-

Decision Rule:- If z is less then -1.96 or greater then +1.96, reject the Null Hypothesis with 95% C.I.

Here z = -5.05 ∴ less then -1.96, so reject the Null hypothesis [There is some fault in the machine]

-1.96 and +1.96 are called Decision Boundaries.

$\boxed{\text{NOTE}}$:-

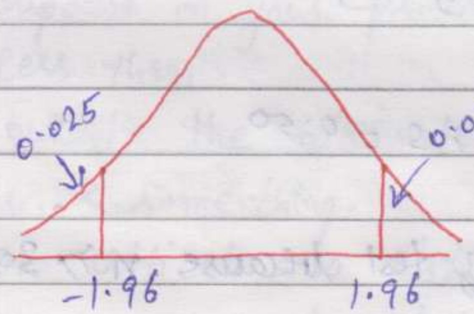
Conditions For Z test :-
① We Know the population standard deviation.
② We do not Know the population standard deviation but our sample is large $n \geq 30$.

Conditions For t test :-
① We don't Know the population Standard Deviation.
② our sample size is small $n < 30$
③ Sample standard deviation is given.

* Another Question :- A complain was registered, the boys in a government school are underfed. Average weight of the boys of age 10 is 32 Kgs with standard deviation = 9 kg. A sample of 25 boys were selected from the government school and the average weight was found to be 29.5 kg? with Confidence Interval C.I. 95%, check it is true or False.

Ans. Null Hypothesis $H_0 : \mu = 32$
Alternate Hypothesis $H_1 : \mu \neq 32$

$\mu = 32$, $\sigma = 9$, $n = 25$, $\bar{x} = 29.5$, C.I = 95%
$\therefore$ S.V = 0.05



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

$$= \frac{-2.5 \times 5}{9} = -1.39$$

Conclusion:-

Z = -1.39 which is between the decision boundaries -1.96 and +1.96, therefore we have to agree with Null Hypothesis.

$\therefore$ At 95% confidence Interval, we fail to reject the null Hypothesis, the boys are fed well.

★ One Tail And Two Tail Test:-

One Tail:- We perform one tail test when the alternative hypothesis states that the parameter is in fact either bigger or smaller than the value specified in the null hypothesis.

⇒ If the following statements are true then we will perform one tail test:-

1. Suppose in your problem statement there are words like less then or more then.
2. In the alternate hypothesis, there is situation like $\mu >$ or $<$ something.
3. Critical area distribution is on the either side of the normal distribution curve [Left or Right]
4. Rejetion area is on the left side or Right side of the normal distribution curve.

★ Question No:-3:- A factory manufactures cars with a warranty of 5 years or more on the engine and transmission. An engineer believes that the engine or transmission will malfunction in less then 5 years. He tests a sample of 40 cars and find the average time to be 4.8 years with a standard deviation of 0.50.

1. State the null hypothesis and alternate hypothesis.
2. At a 2% significance value, is there enough evidence to support the idea that the warranty should be revised.

Ans:- Here we have to perform one tail test, becaue alternate hypothesis state that the parameter is less then null hypothesis.

Null Hypothesis (H_0) = $\mu \geq 5$
Alternate Hypothesis (H_1): $\mu < 5$

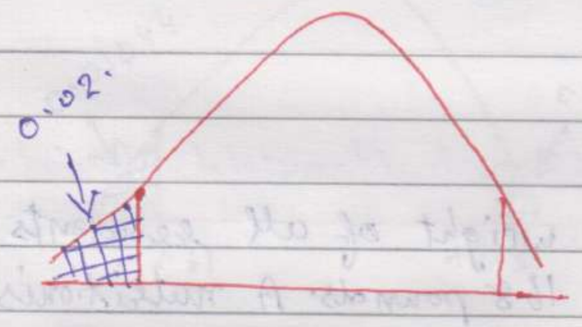
Now:- $\mu = 5$, $n = 40$, $\bar{x} = 4.8$, $S = 0.50$

We have to perform the z test because $n \geq 30$

$$Z\,test = \frac{\bar{x} - \mu}{S/\sqrt{n}} = \frac{4.8 - 5}{0.50/\sqrt{40}} = \frac{(-0.2) \times 6.324}{0.5}$$
$$= -2.5$$

$$\boxed{Z = -2.5}$$

Decision Boundaries:-



Significance value = 0.02

$Z_{\alpha/2}$ =

As it is a one tail test
$Z_\alpha = Z_{0.02} = -2.05$

Decision Boundaries are -2.05 and $+2.05$

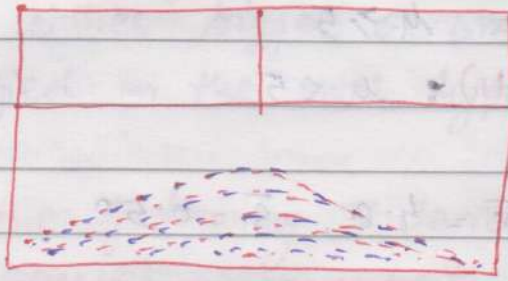
Conclusion:-

$-2.5 < -2.05$ ∴ Rejet the Null Hypothesis

Warranty needs to be revised.

★ **P-value**

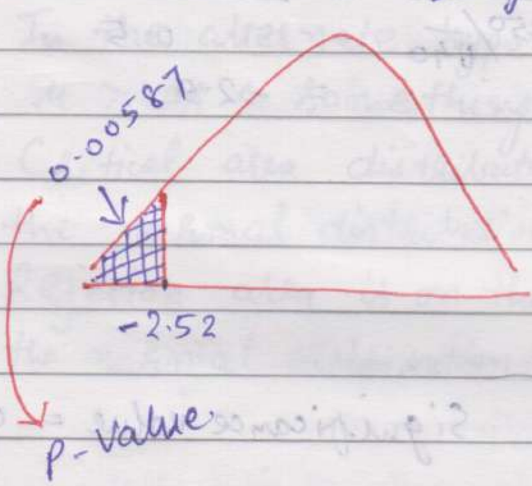
With the help of Zscore, whatever area you r getting under the curve is your p value.



From the above example

Z test = -2.52

$\alpha = 0.02$ [significance value]



P value < α [Yes]

Reject the null hypothesis.

★ **Question 4:-** The average weight of all residents in a town xyz is 16.8 pounds. A nutritionist believes the true mean to be different. She measured the weight of 36 individuals and found the mean to be 169.5 pounds with a standard deviation of 3.9

(a) Null & Alternate Hypothesis

(b) 95% Confidence Interval. Is there enough evidence to discard the null hypothesis?

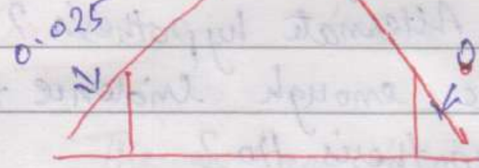
Ans:- Null Hypothesis (H0): $\mu = 16.8$

Alternate Hypothesis (H1): $\mu \neq 168$

Here $\mu = 168$, $n = 36$, $\bar{x} = 169.5$, $S = 3.9$, $C.I = 0.95$

Step 2 :- $C.I = 0.95$ $\alpha = 0.05$

Step 3:- It is a two tail test.



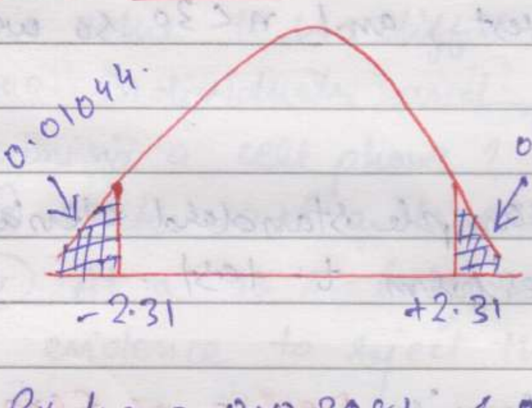
Decision Boundaries are -1.96 and $+1.96$ acording to Z table.

Step:-4:- Z-score $= \dfrac{\bar{x} - \mu}{S/\sqrt{n}} = \dfrac{169.5 - 168}{3.9/\sqrt{36}} = \boxed{2.31}$

Conclusion:-

Z score $2.31 > 1.96$ {Reject the Null Hypothesis}

Now P value :-



$0.01044 + 0.01044 = $ P-value.

$\therefore$ Pvalue $= 0.02081$

Pvalue $= 0.02081 < 0.05 = \alpha$ (Significance value)

When Pvalue is less then S.V {Reject the Null Hypothesis}

Note:- If Pvalue $> \alpha$ then we have to agree the null hypothesis.