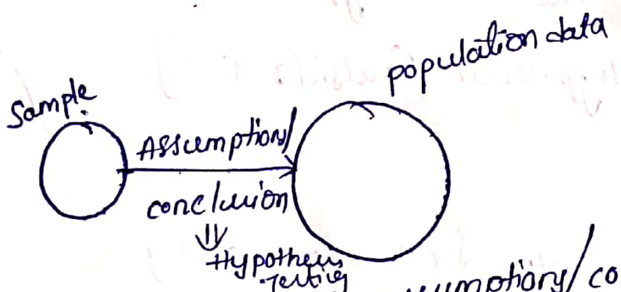
Inferential Statistics:

Agenda:

→ Hypothesis Testing
→ p-value
→ Confidence Interval
→ Significance value.

Inferential stats



→ we are making some assumptions/conclusions on sample data population data based on the sample data.

→ To validate the assumptions/conclusions we use the hypothesis Testing.

Steps of hypothesis Testing:

① Null Hypothesis

② Alternate Hypothesis.

③ Perform Experiments

Ex: Experiment

[coin is fair or not]

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

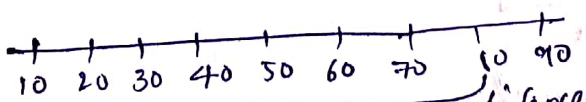
- Then

1). Null hypothesis = coin is fair

2). Alternate hypothesis : coin is not fair.

3) Perform Experiment

CI = (20-80)
⇓
coin is fair



CI → confidence Interval

If coin is tossed 100 times

→ 50 times head ⇒ fair

→ 60 times Head ⇒ fair

→ 70 times Head ⇒ fair

→ 80 times Head ⇒ not fair

CI = [20 − 80]

⇓

~~if co~~

coin is fair

⇒ we fail to Reject the Null Hypothesis [within C.I]

→ we reject the Null Hypothesis [outside C.I]

→ conclusion

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.

⇓

judge → conclusions

Conclusions

confidence Interval (CI) ⇒ Given by domain expert

$CI = \boxed{95\,\%}$

Significance value =

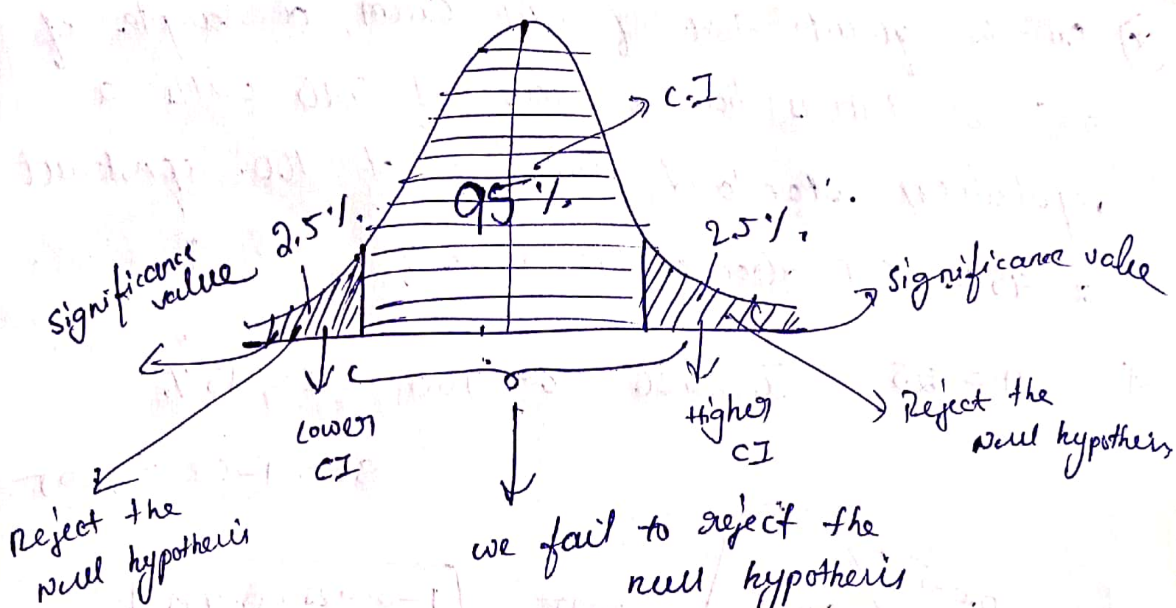
~~sif~~ significance value = 1 − C.I

$SV = 1 - 0.95 = 0.05$

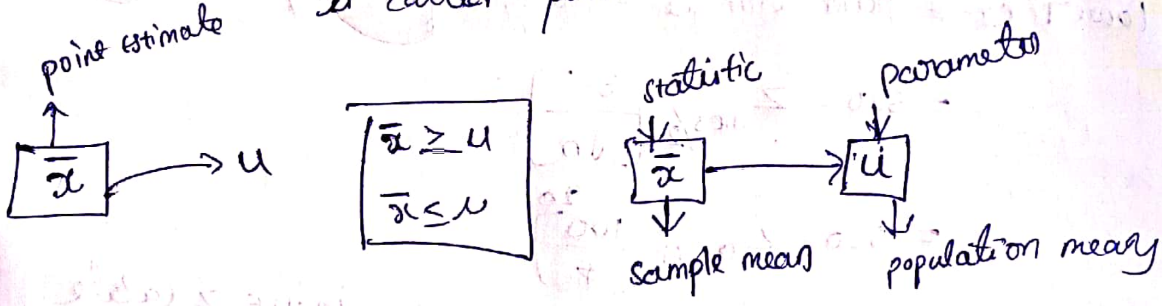
$$\boxed{SV = 1 - C.I}$$

SV → significance value

CI → cofidence Interval



point Estimate : The value of any statistic that estimates the value of a parameter is called point Estimate



$\boxed{\text{point Estimate}} \pm \boxed{\text{margin of Error}} = \boxed{\text{parameter} \Rightarrow \text{population mean}}$

Lower C.I : point estimate $-$ margin of error

Higher C.I : point Estimate $+$ margin of error

margin of Error $= Z_{\alpha/2} \left[\frac{\sigma}{\sqrt{n}} \right]$ $\sigma \rightarrow$ population s.d $\rightarrow$ standard error $\alpha =$ significance value

$\alpha \Rightarrow$ significance value.

$\sigma \Rightarrow$ population standard deviation

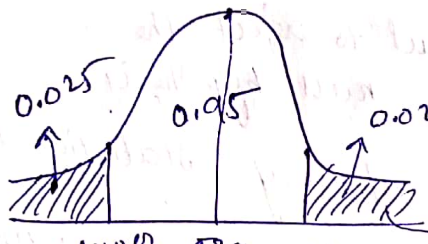
$\left[\frac{\sigma}{\sqrt{n}} \right] \Rightarrow$ Standard Error

Scanned with CamScanner

(*) 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:- $n = 25$ $\bar{x} = 520$ $\sigma = 100$ $CI = 95\%$

$SV = 1 - CI = 0.05$



$\boxed{1 - 0.025 = 0.975}$

$\alpha = 0.025$

lower CI = point estimate − margin of error

$= 520 - Z_{0.05/2}\left[\frac{\sigma}{\sqrt{n}}\right]$

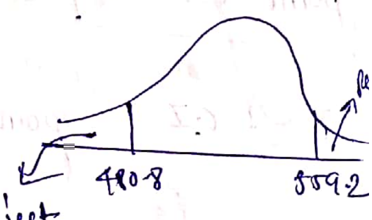
$= 520 - Z_{0.025}\left[\frac{100}{5}\right]$ (20)

see in the z-table

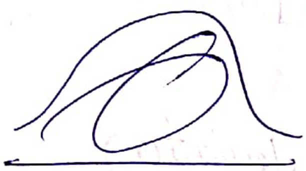
$= 520 - 1.96 \times 20.$

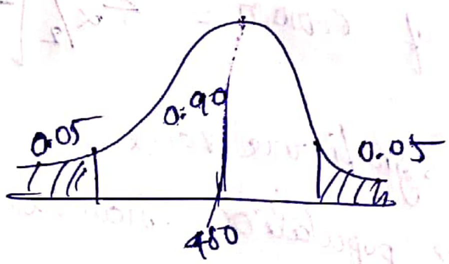
$= 480.8$

Higher C.I = $520 + 1.96 \times 20 = 559.2$



(*) $\bar{x} = 480$ $\sigma = 85$ $n = 25$ $CI = 90\%$ $SV = 1 - 0.90 = 0.10$





lower CI $= 480 - Z_{0.10/2}\left(\frac{85}{5}\right)$

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

$= 480 - 1.64(17)$

$= 480 - 27.8 = 452.12$

Higher CI $= 480 + 1.64(17)$

$= 480 + 27.8$

$= 507.8$

$CI = [452.12 \longleftrightarrow 507.8]$

* 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% CI about the mean?

sol: $\bar{x} = 520$ $\quad S = 80$ $\quad n = 25$ $\quad CI = 95\%$ $\quad SV = 1 - 0.95 = 0.05$

~~for~~ $\bar{x} \pm t_{\alpha/2}\left(\frac{S}{\sqrt{n}}\right)$

for test
we need to find
degree of freedom $= \boxed{n-1}$
$= 25 - 1 = 24$

lower $CI = 520 - t_{0.05/2}\left(\frac{\overset{16}{\cancel{80}}}{\cancel{5}}\right)$ → See t table for this value

$= 520 - 2.064 * 16$

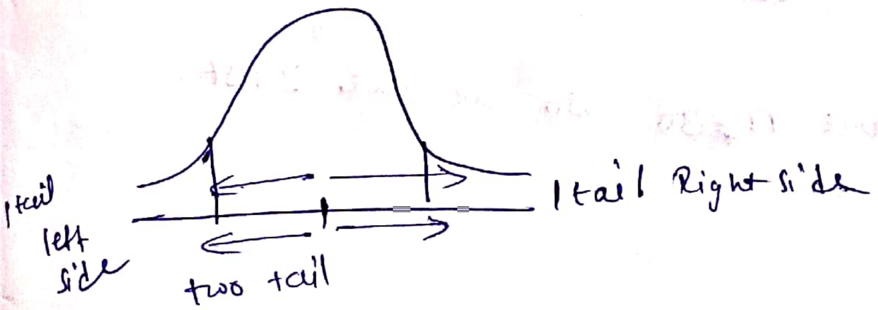
$= 486.976$

Higher $CI = 520 + 2.064 * 16$

$= 553.024$

$CI = [486.976 \longleftrightarrow 553.024]$

* 1 tail and 2 tail Test



Hypothesis Testing problems

① 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 samples, he measures the average amount dispensed by the machine to be 78 ml with a standard deviation of 2.5.

(a) State Null & Alternate Hypothesis

(b) At 95% CI, is there enough evidence to support machine is working properly or not.

Sol:

$\mu = 80ml \quad n = 40 \quad \bar{x} = 78ml \quad s = 2.5$

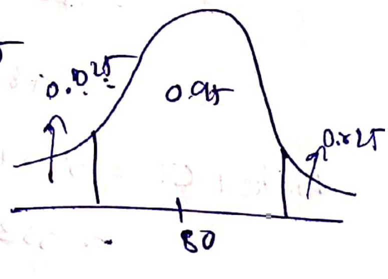
Step 1:

Null Hypothesis : $\mu = 80$

Alternate Hypothesis : $\mu \neq 80$

Step-2:

$C.I = 0.95 \quad S.V(\alpha) = 1 - 0.95 = 0.05$



Step 3:

$n = 40$

$s = 2.5$

① $n \geq 30$ or population sd $\rightarrow$ use z test

② $n < 30$ and sample std $\rightarrow$ use t test

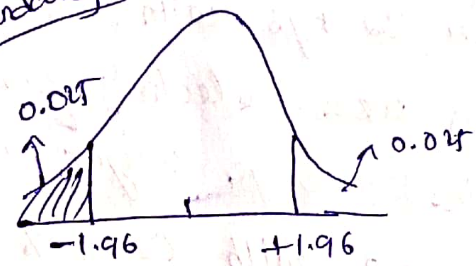
So,

Here we have $n \geq 30$ so we use z test.

Z-test

lets perform the experiment

Decision boundary



$1 - 0.025 = 0.975$

* calculate test statistics (Z-test)

$$Z = \frac{\bar{x} - \mu}{S/\sqrt{n}} = \frac{78 - 80}{\frac{2.5}{\sqrt{40}}} = -5.05$$

* conclusions

Decision rule: If $Z = -5.05$ is less than -1.96 or greater than $+1.96$, Reject the Null hypothesis with 95% CI

Reject the Null Hypothesis { there is some fault in the machine.

* Conditions for Z-test:

① we know the population sd. OR

② we do not know the population sd but our sample is large $n \geq 30$.

* Conditions for t-test:

① we don't know the population sd.

② our sample size is small $n < 30$.

③ sample sd is given.

Scanned with CamScanner

(2) A complaint was registered, the Boys in a government School are underfed. Average weight of the Boys of age 10 is 32 kgs with S.D = 9 kgs. A Sample of 25 Boys were Selected from the government School and the average weight, was found to be 29.5 kgs ? with CI = 95%. Check it is True or False.

Sol: $n = 25$ $\mu = 32$ $\sigma = 9$ $\bar{x} = 29.5$ $CI = 0.95$

$SV = 1 - CI = 0.05$

Step 1:

(1) Null hypothesis: $\mu = 32$

(2) Alternate hypothesis: $\mu \neq 32$

Step-2:

$CI = 0.95$

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

Step-3:

Z- test

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

Conclusions: $-1.39 > -1.96$. Accept the null hypothesis 95% CI

we fail to reject the null hypothesis

The Boys are fed well.

Scanned with CamScanner

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 than 5 years. He tests a sample of 40 cars and finds the average time to be 4.8 years with a standard deviation of 0.50.

ⓐ State the null & alternate hypothesis.

ⓑ At a 2% significance level, is there enough evidence to support the idea that the warranty should be revised?

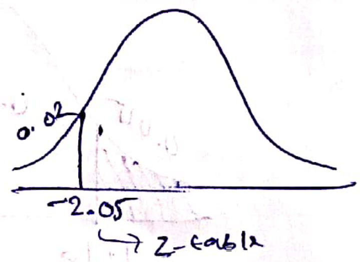
sol: $n = 40 \quad \mu = 5 \quad \bar{x} = 4.8 \quad s = 0.50$

Step 1:

① Null Hypothesis : $\mu \geq 5$

② Alternate hypothesis : $\mu < 5$

step 2:



$\alpha = 0.02 \qquad CI = 0.98$

step 3: Z-test

$$Z\text{-score} = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{4.8 - 5}{0.50/\sqrt{40}} = \frac{-0.2}{0.079} = -2.531$$

conclusions:

$-2.53 < -2.05$, Reject the null Hypothesis

warranty needs to be revised.

④ In the population the average IQ is 100 with a standard deviation of 15. A team of Scientists want to test a new medication to see if it has a +ve or -ve effect, or no effect at all.
A Sample of 30 participants who have taken the medication has a mean of 140. Did the medication affect Intelligence? {95%}

Sol:- $\mu = 100$ $\sigma = 15$ $n = 30$ $\bar{x} = 140$ $CI = 95\%$

Step-1:

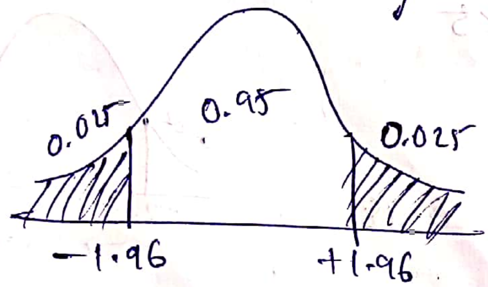
① Null Hypothesis: medication has an effect

② Null Hypothesis: medication doesn't affect

Step-2:

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

Step-3 Decision boundary



Step 4- calculate test statistics (z-test)

$$Z\text{ score} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{140 - 100}{15/\sqrt{30}} = \frac{40}{2.738} = 14.619$$

conclusion:

14.619 > 1.96, Reject the Null hypothesis.

medication doesn't affect any intelligent.