Day - 5

Statistics

Inferential statistics:

① Hypothesis Testing

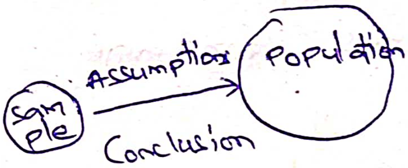
② P- value

③ Confidence Interval

④ significance value

Hypothesis testing: is to test whether the null hypothesis can be rejected or approved

Inferential stat:



⇓

Hypothesis Testing

→ Experiments

steps of hypothesis Testing

① Null Hypothesis: Coin is fair

(Coin is fair or not)

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

② Alternative Hypothesis: Coin is not fair

③ Perform Experiments:

100 times

60 40
70 30
80 20

⇓

50 times Head ⇒ fair

60 Head

70 times ⇒ Domain Expert

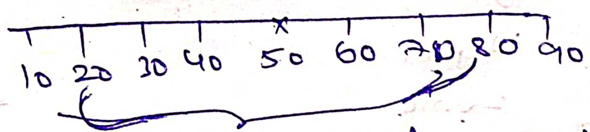
⇓

Confidence Interval

10 → Null Hypothesis is rejected

→ Alternative Hypothesis is accepted

C.I = [20-80] ⇒ coin is fair



C.I ⇒ Confidence Interval

↳ 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 Altenative Hypothesis: Person is criminal

③ Experment/proof: DNA, fingerprint, weapons, eye witnes, footage

⇓

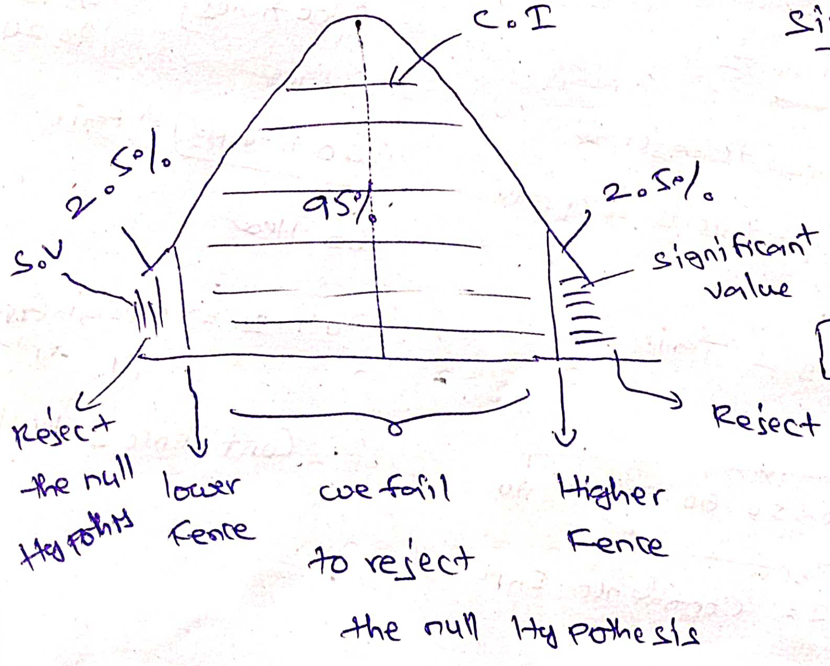
Judge

Conclusions:

Confidence Interval (C.I):

$\boxed{95\%}$ ⇒ Domain Expert

$\boxed{1-CI \quad S.V = (1-C.I)}$



Significence value:

C.I = 95%

SV = 1 − 0.95

$\boxed{S.V = 0.05}$

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

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

statistics $\qquad$ Parameter

$\uparrow \qquad\qquad \uparrow$

Point Estimate

$\Uparrow$

$\boxed{\bar{x}} \longrightarrow \mu$

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

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

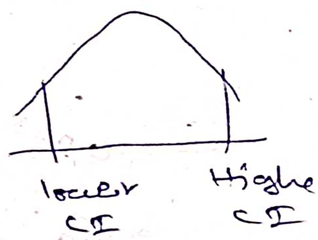
$\uparrow$

sample mean

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

lower C.I : Point Estimate − margin of Error

Higher C.I : Point Estimate + margin of Error



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

$\sigma \rightarrow$ Population sd

$\Rightarrow$ standard Error

α = signification value

* Problems:

\# 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.

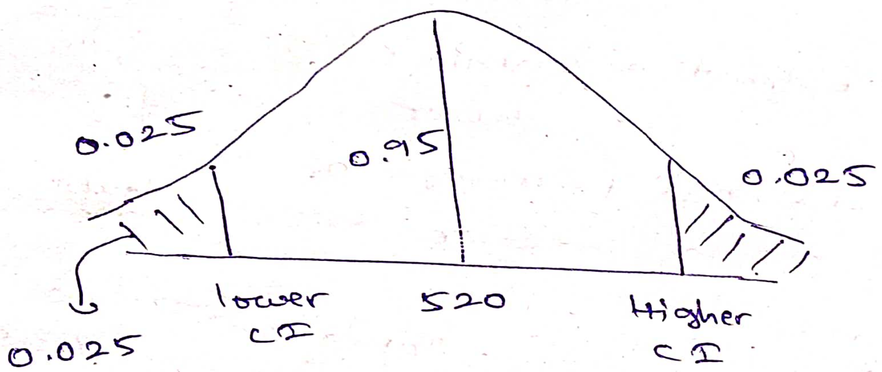
Ans:

$n = 25 \quad \bar{x} = 520 \quad \sigma = 100 \quad CI = 95\%$

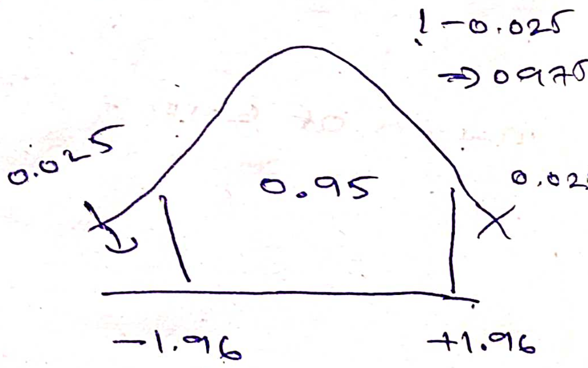
$S.V = 1 - CI = 0.05$
$= 1 - 0.95$

$5/2 = 2.5 \Rightarrow 0.025$

$1 - 0.025 \Rightarrow 0.975$



$\boxed{\alpha = 0.05}$



lower CI = Point Estimator − margin of Error

$\Rightarrow 520 - Z_{0.05/2} \dfrac{\sigma}{\sqrt{n}}$

$\Rightarrow 520 - Z_{0.025} \dfrac{\cancel{100}\,20}{\cancel{5}\,1}$

$\Rightarrow 520 - 1.96 \times 20 = 480.8$

$\Rightarrow 480.8$

Higher CI $= 520 + 1.96 * 20$

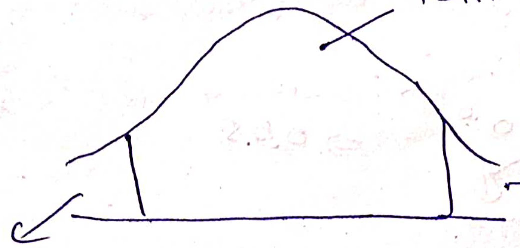
$= 559.2$

$CI = 0.95$

$SV = 1 - 0.95 = 0.05$

$Z_{0.05/2} = \boxed{Z_{0.025}}$

* use Z table



* $\bar{x} = 480 \quad \sigma = 85 \quad n = 25 \quad CI = 90\%$

Significance value $= 1 - CI$
$= 1.090$
$= \boxed{0.10}$

lower C.I $= 480 - Z_{0.10/2} \left[\frac{85}{5}\right]$

$= 480 - Z_{0.05} \left[\frac{85}{5}\right]$

$= 480 - 1.64 \, [17]$

$= 480 - 27.8 = 452.12$

Higher C.I
$= 480 + 1.64 \, (17)$
$= 507.8$

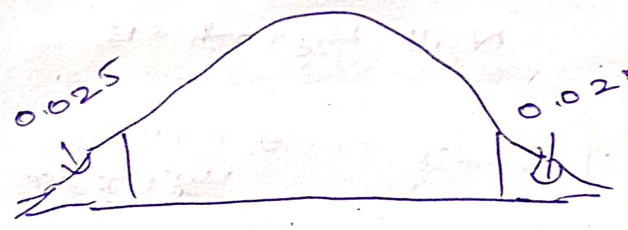
② On the quant test of CAT exam, a sample of 25 test takers has a mean of 520, with a sample standard deviation of 80 Construct 95% C.I about the mean

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

$n = 25$

t-test

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



$\Rightarrow 0.05$

t-test

Degree of freedom

$= n-1$

$\Rightarrow 25-1 = \boxed{24}$

lower C.I = Point Estimate − margin of Error

$= 520 - t_{0.05/2}\left(\frac{\overset{16}{\cancel{80}}}{\cancel{5}}\right)$ $\boxed{t_{0.025}}$

$= 520 - 2.064 * 16$

lower C.I $= 486.976$

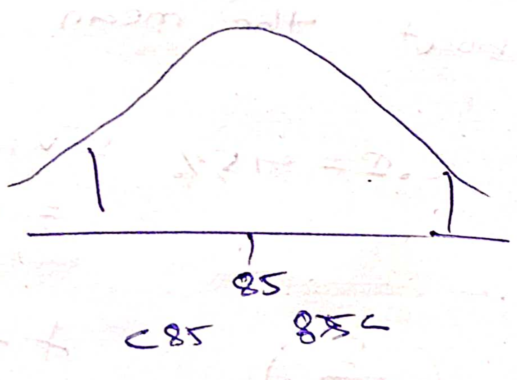
* $\boxed{\text{use T-table}}$

~~lower~~ higher C.I $= 553.024$

* 1 tail and 2 tail Test :-

* Colleges in the 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?

Ans:- Two tail



↓
greater than or less than 85%.

* Hypothesis testing:

* A factor has a machine that fails 80ml of baby medicenes in a bottle an employe belives the average amount of baby medicience is not 80ml. using 40 sample, he measures the average amount dispersed by the machine to be 78ml with a standard deviation of 2.5

a) State Null of Alternatie Hypothesis

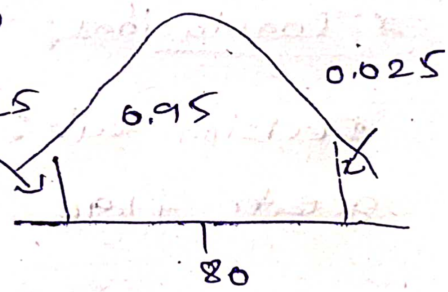
b) at 95% C.I, is there enough evidance to support machine is working properly or not

Step 1

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

Ans: Null Hypothesis $\mu = 80$

Alternative Hypothesis $\mu \neq 80$



Step 2: C.I = 0.95 S.V(α) = 1 − 0.95
= 0.05

Step 3:

n = 40

s = 2.5

① $\boxed{n \geq 30 \text{ or population sd}}$ } z-test

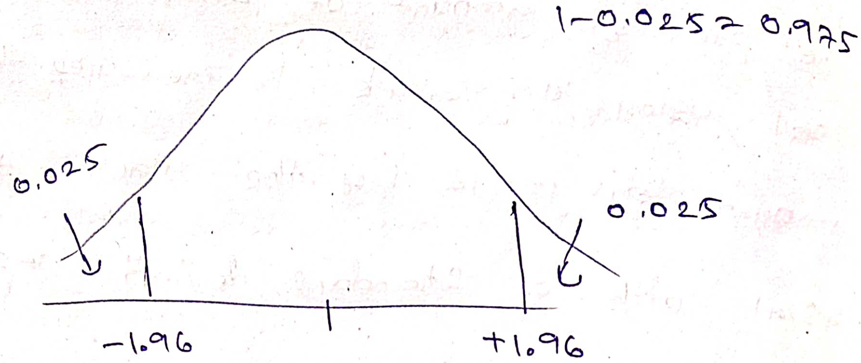
② $\boxed{n < 30 \text{ and sample sd}}$ } t-test

z-test

Z-test

let perform the Experment

Decision Boundary



* Calculate test statistics (Z-test)

$$Z = \frac{\bar{x} - \mu}{\boxed{s/\sqrt{n}}} \rightarrow \text{standard Error} \qquad = \frac{78-80}{\frac{2.5}{\sqrt{40}}} = -5.05$$

* Conclusions:

Decision Rule: If $z = -5.05$ is less than -1.96 or greater $+1.96$ Reject the Null Hypothesis with 95% CI

Reject the null Hypothesis } There is some fault in the machine

* A complaint was registered, the boys in a Government School are under fed, Average weight of the boys of age 10 is 32 kgs with S.D=9kgs

A sample of 25 boys were selected from the Goverement school and the average weight was found to be 29.5 kgs? with CI =95%

check if is True or False

Ans: Conditions for z-test $n=25$, $\mu=32$ $\sigma=9$ $\bar{x}=29.5$

① we know the population sd (or)

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

Conditions from T test

① we do not know the population std

② our sample size is small $n<30$

③ Sample std is given

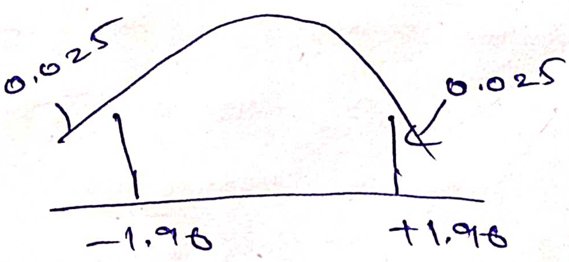
Step 1

① $H_0 : \mu = 32$

$H_1 : \mu \neq 32$

② C.I $= 0.95$ $\alpha = 1-0.95 = 0.05$

③ z test



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

Conclusion:- $= -1.39 > -1.96$ Accept the null Hy-pothesis 95% C.I we fail to reject the Null Hypothesis the Boys are ~~fed~~ well

Day – 6

* A factory manufactures cars with a warranty of 5 years or more on the engine and transmission. An engineer belives that the engine or transmission will malfunction is less than 5 years. He tests a sample of 40 cars and finds the average time to the 4.8 years with a standard deviation of 0.50

(1) state that null & alternative hypothesis

(2) At a 2% significance level is there enough evidance to support the idea that the warranty should be recived?

Ans: $n=40$ $\bar{x} = 4.8$ years $s = 0.50$

Step 1:

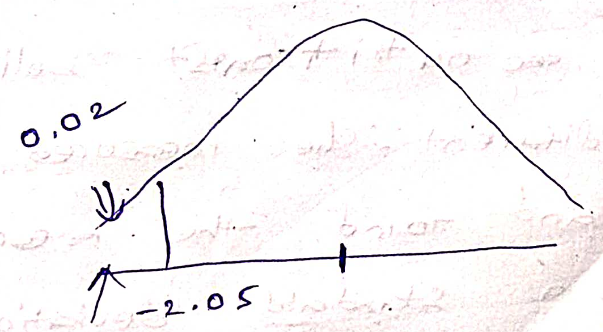
$H_0 : \mu \geq 5$ {Null hypothesis}

$H_1 : \mu < 5$ {Alternative hypothesis}

Step 2:

$\alpha = 0.02$ $C.I = 1 - 0.02 = 0.98 = 98\%$

Step 3:



Reject null Hypothesis.

Step 4:

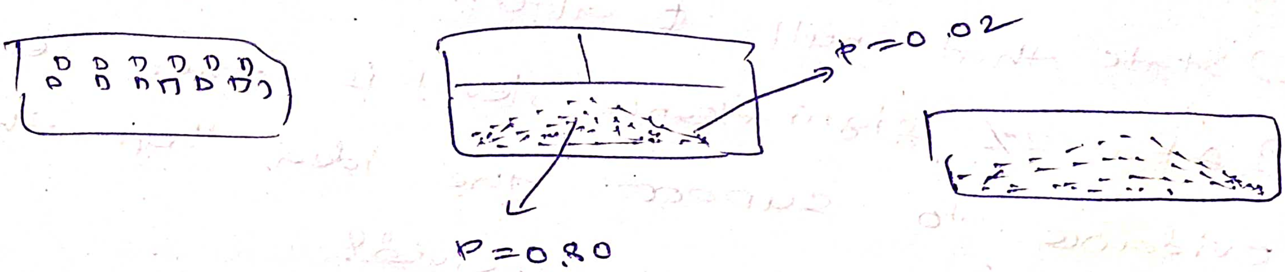
$$z\text{-score} = \frac{\bar{x}-\mu}{s/\sqrt{n}} = \frac{4.8-5}{0.50/\sqrt{40}} = -2.5298$$

Conclusion: $-2.52 < -2.05$ Reject the null Hypothesis

Warranty need to be revised

P_value: {significance value}

Keyword touch



$\alpha = 0.02$

Pvalue < α [Yes]

Reject the Null Hypothesis

p_value $\boxed{-2.5298}$

* the average weight of all residents in a town XYZ is 168 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 39

a) Null & Alternate Hypothesis

b) 95% Is there enough evidence to discard the null hypothesis?

Ans: $\bar{x} = 169.5$ $\quad s = 3.9$ $\quad n = 36$ $\quad \mu = 168$

$C.I = 0.95$

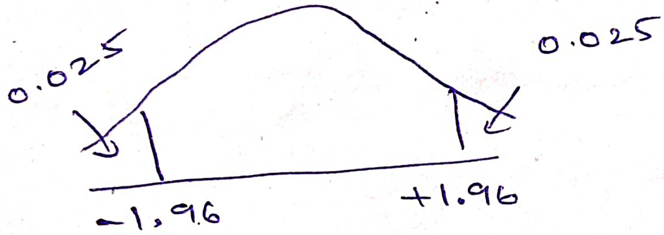
Step 1:

$H_0 \quad \mu = 168$

$H_1 \quad \mu \neq 168$

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

Step 3:



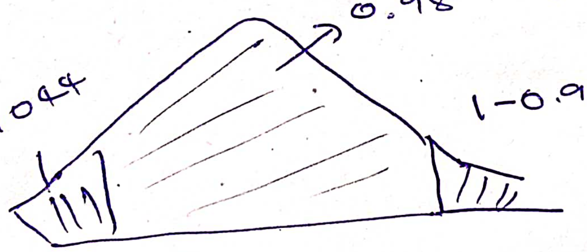
Step 4:

$$z\ score = \frac{169.5 - 168}{3.9/\sqrt{36}} = \approx 2.31 \Rightarrow 2.3076$$

$2.31 > 1.96$ {Reject the null Hypothesis}

2 Tail test

P-value:



$1 - 0.98956 = 0.01044$

$P.\ value = 0.01044 + 0.01044 = 0.02088$

$0.02088 < \underset{\alpha}{0.05}$ {Reject the Null Hypothesis}

* chi square test

* chi square test claims about population proportion. It is a non parametric test that is performed on categorical data ⟹ ordinal data, Nominal data

* In the 2000 u.s census the age of individuals in a small town found to be the following

<18	18-35	>35
20%	30%	50%

In 2010, ages of $n=500$ individuals were sampled. Below are the results

<18	18-25	>35
121	288	91

using $\alpha=0.05$ would you conclude the population distribution of ages has changed in the last 10 years?

Ans:

	<18	18-35	>35
Expected	20%	30%	50%

$n=500$

	<18	18-35	>35
Observed	121	288	91
Expected	100	150	250

Step 1:- Null Hypothesis

Ho : the data meets the expected distribution

H_1: The data does not meet the expected distribution

Step 2:- $\alpha = 0.05$ C.I = 95%

Step 3:- Degree of freedom {Categories}

$df = C - 1 = 3 - 1 = 2$ $\alpha = 0.05$

↳ No. of categories.

Step 4:- Decision Boundary = $\boxed{15.991}$ {chi square table}

Step 5:- Chi Square Test Statistics

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(121-100)^2}{100} + \frac{(288-150)^2}{150} + \frac{(91-250)^2}{250}$$

$$\boxed{\chi^2 = 232.494}$$

Conclusion:-

$\chi^2 > 5.99$ Reject Ho