Day 5 :-

1. Covariance, pearson and Spearman Ran[k]

   formula. $\rho_{x,y} = \frac{Cov(x,}{\sigma_x \cdot \sigma}$

2. Inferential Stats

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

* Pearson Correlation Coefficient [-1 to

If X and y has relation of 0.8 and X and Z has relation of 0.9 it is called highly Co-related.

$$X \xleftrightarrow{0.8} Y \leftrightarrow Z \qquad X \xrightarrow{0.9} Z$$

0 = no Corr[elation]

+1 = more pos[itive] Correlation

-1 = more nega[tive] Correlated.

$\boxed{X\ Y = 0.8}$ ✓

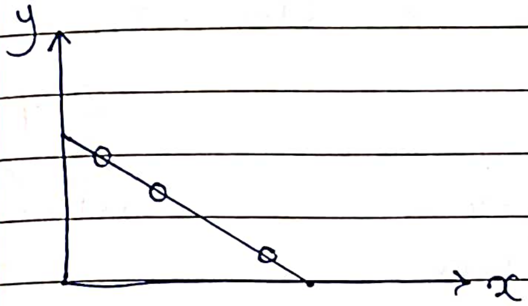
$\boxed{X\ Z = 0.9}$ ✓

↳ 1 +ve Correlated.

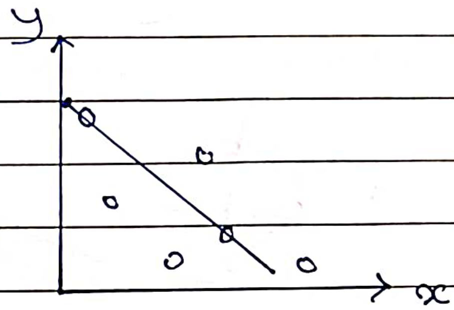
Between {0 To 1} {-1 to 0}

$$\boxed{\begin{array}{c} x\uparrow \text{ and } Y\downarrow \\ X\downarrow \text{ and } Y\uparrow \end{array}}$$

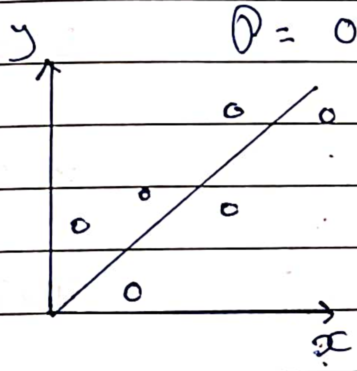
$\rho = -1$



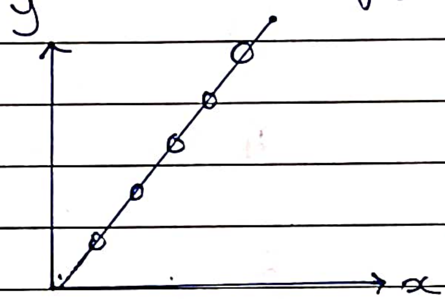
$\rho = -1 < 0$ {Between -1 to 0}



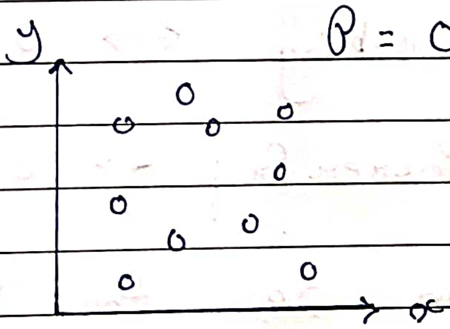
$\rho = 0$ to 1



$\rho = +1$



$\rho = 0$



Examples of Scatter diagrams with values of Correlation Coefficient.

* Why we are learning Covariance, ~~pkts~~ Pearson and Spearman Rank Correlation.

→ Lets Say we have 3 elements.

Machine Learning model.

Output fe

Age	Working Exp	Carrer Gap	Salary
24	2	1	45000
32	10	3	50000
25	4	0	35000
26.	5	0	40000

a] Age ⟷ working ⟷ 0.9

b] Age ⟷ Carrer Gap ⟷ 0.9

As we can see, the Corelation between this two are Same, So in machine learning we can reject one of this and train the model with only two features. As this technique is features Selection.

Feature Selection (Correlated)

{Age ↔ Working Exp = 0.9}

{Age ↑ Work Exp ↑}
{Age ↓ Work Exp ↓}

{Age ↔ Carrer Gap = 0.9}

By using only One feature we can train module.

* Disadvantage

(i) Pearson Correlation works well with Linear data (Straight Line).

Spearman Rank Correlation

X	Y	Rank x	Rank y
1	2	4	4
3	4	3	3
7	5	5	2
0	7	2	1
8	1	1	5

Formula Spearman

$$r_s = \frac{Cov(R(x), R(Y))}{\sigma R(x), \sigma R(Y)}$$

Spearman Rank is used with non-Linear dataset

Page No.
Date

* Spearman Rank Correlation

It only apply on Ranks for non-linear datasets with an individual

Formula :

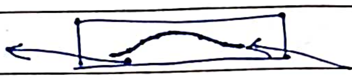
$$R_s = \frac{Cov(R(x), R(y))}{\sigma R(x) \cdot \sigma R(y)}$$

X	Y	(R)x	(R)y
3	6	2	7
8	5	3	5
0	7	1	0
10	3	5	2
9	4	4	3
17	2	6	1
25	1	7	6

Inferential Statistics

(1) P value. {Hypothesis Testing}

P-value = 0.8 — Out of all hundred touches Iam using this are Just 2 times.

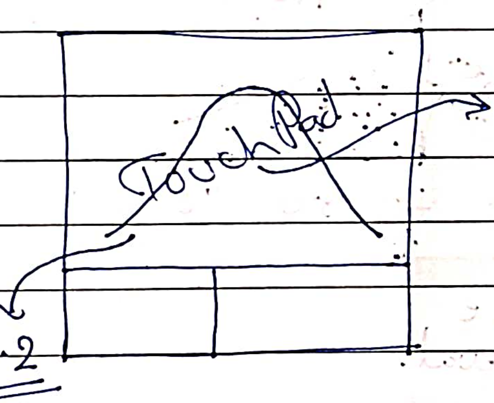


Space bar in Keyboard.

Gaussian dist.

*. P value is the porbability of the null hypothesis is True?

P - value.



P = 0.8 — Out of all 100 the Touch over This are is 80%.

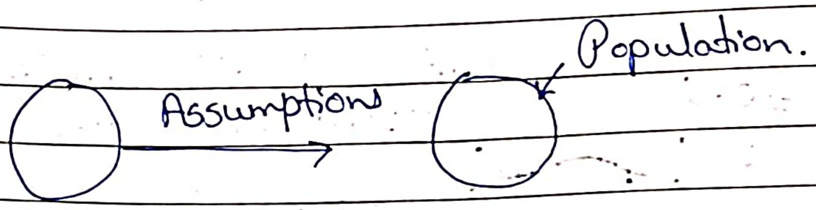
P = 0.2

ecial

- Point Estimate : The value of any Statistics that estimates the Value of a parameters is called point estimate.

Point estimate $\boxed{\bar{x}} \xrightarrow{\text{Estimate}} \boxed{\mu} \longleftarrow$ Parameter.

* Inferential Statistics we make an assumption over regarding population data.



* Steps of Hypothesis

(i) Null Hypothesis → Coin is fair (H_0)

(2) Alternate Hypothesis → Coin is not fair (H_1)

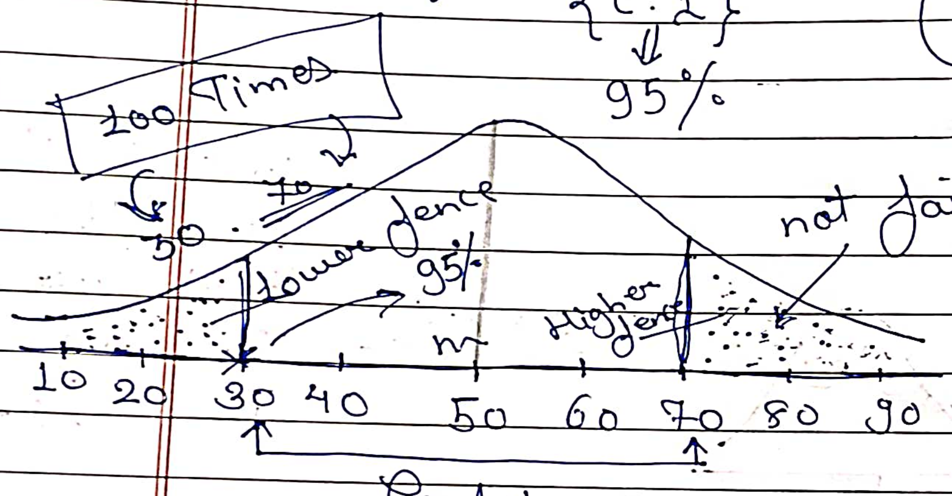
(3) Performe Experiment.

{C.I} ⇓ 95%

(1) → Coin is fair or not

⇓

Experiment [Toss]



10 :- Null Hypothesis is rejected

30 : Null Hypothesis is accepted

Significance Value = 0.05 ⇑ 95%

Significance Value = 1 - C.I

= 1 - 95% = 0.05

* Point estimate

The value of any statistic that estimates the value of a parameter is point estimate.

Point estimate → $\boxed{\bar{x}}$ (Sample mean) ←estimate→ $\boxed{\mu}$ ← Parameter

μ : Population mean

20, 45, 30   40

$\boxed{\overset{+}{20}}$ $\boxed{\overset{-}{5}}$ $\boxed{\overset{+}{10}}$ → margin of error.

Point Estimate $\pm$ margin of error = Param

↓

How much more or less that the parameter.

Lower fence = Point estimate − margin of error

Higher fence = Point estimate + margin of error.

Q. In the Quant test of CAT exam, the population standard deviation is known to be 100. A sample of 25 test takers has a mean of 520. Construct a 95% C.I about mean?

→ Ans)

$\sigma = 100$

$n = 25$

$m = 520$

Population S.D ∴ Z-table

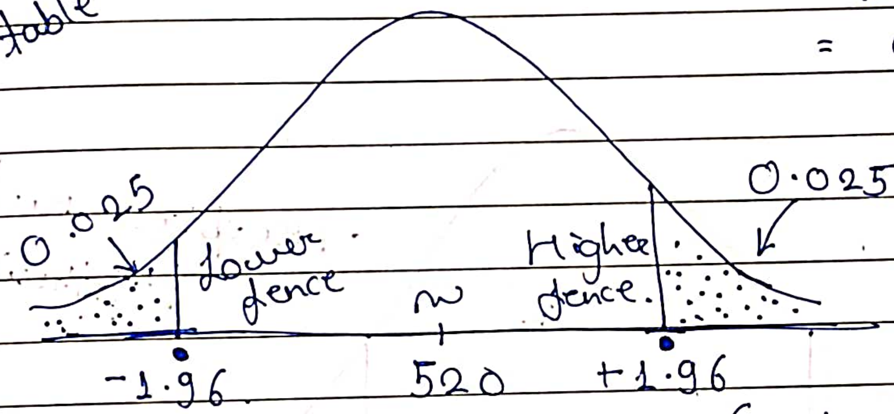
$CI = 95\%$

α = significance value

∴ 1 − C.I

∴ 1 − 0.95

= 0.05 (half divide)



∴ 1 − 0.025

= 0.975

(refer Z table)

Point estimate + margin of error.

$$\bar{x} + Z_{\alpha/2} \boxed{\frac{\sigma}{\sqrt{n}}} \Rightarrow \text{standard error.}$$

lower fence. $= \bar{x} - Z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}}$

$Z = \dfrac{0.05}{2} = Z_{0.025}$

$= 520 - 1.96 \dfrac{100}{\sqrt{25}}$

$= 520 - 1.96 \times 20$

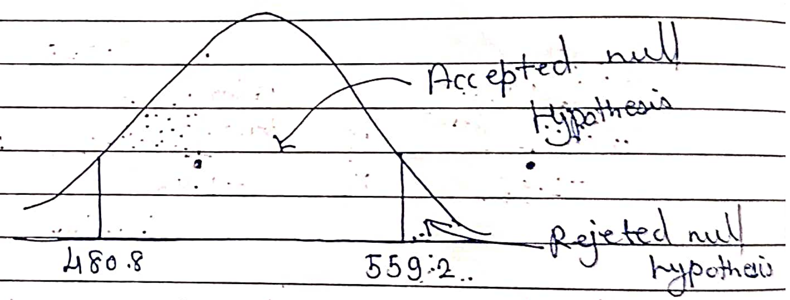
$= \underline{\underline{480.8}}$

* Higher fence

$$\bar{x} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

$$\therefore 520 + 1.96 \frac{100}{\sqrt{25}}$$

$$\therefore 520 + 1.96 \times 20 = 559.2$$



* On the quant Test of CAT Exam A Sample of 25 test taker has mean of 520 with a Sample Standard deviation of 80. Construct 95% about the mean.

$t_{\alpha/2} = \underline{2.064}$

→ $x = 520$

$S = 80$

$C.I. = 95\%$

$\alpha = 0.05$

Sample Standard deviation [T - test]

Degree of freedom $= n - 1 = 25 - 1 = \boxed{24}$

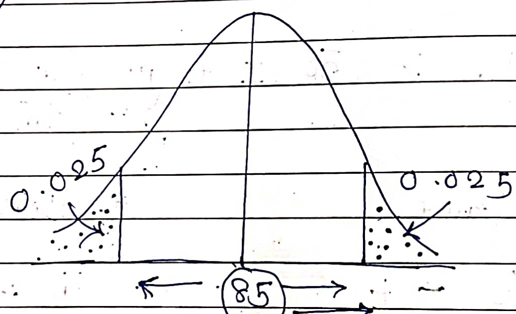
Formula : $\bar{x} \pm t_{\alpha/2} \left(\frac{s}{\sqrt{n}} \right)$

* Lower fence:	* Higher fence:
$\bar{x} - t_{\alpha/2} \left(\frac{s}{\sqrt{n}} \right)$	$\bar{x} + t_{\alpha/2} \left(\frac{s}{\sqrt{n}} \right)$
$\therefore 520 - 2.064 \frac{80}{\sqrt{25}}$	$\therefore 520 + 2.064 \frac{80}{\sqrt{25}}$
$\therefore 520 - 2.064 \left(\frac{80}{5} \right)$	$\therefore 520 + 2.064 \times \frac{80}{5}$
$\therefore \underline{486.976}$	$\therefore \underline{553.022}$

Interview Question

One Tail & 2 Tail.

Q. College in a Town A has a 85% placement rate. A new College was recently openeded and it was found that a sample of 150 student had a placement rate of 88% with a standard deviation of 4%. Does this College has an different placement rate with C.I of 95%



2 Tail   1 tail > 85%

* Condition for Z Test

(i) we know the population S.d, OR
(2) we do not know population S.d but our sample is large $n \geq 30$.

-Test

(1) we know the population variance OR,
(2) our sample is small $n \leq 30$.

Page No.
Date

* Assignment

In Cat exam the population standard deviation is 100 and sample is 25, mean is 520, Construct the Confidence interval of 80 % About the mean ?

→ Significance value

$\alpha = 1 - 0.80$

$= \underline{0.20}$

$\sigma = 100$

$n = 25$

$m = 520$

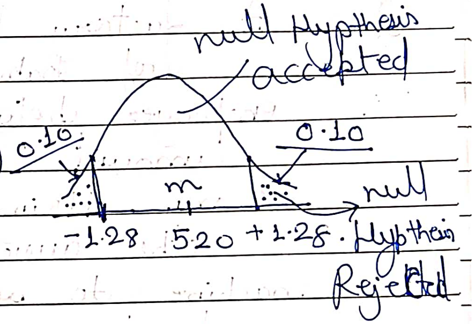
$C.I = 80\%$

$\frac{0.20}{2} = 0.10$

$\therefore 1 - 0.10 = \underline{0.90}$ (z score Table)

$\Downarrow$

$\underline{1.28}$



Lower fence

ula :- $\bar{x} \pm z_{\alpha/2} \left(\frac{\sigma}{\sqrt{n}}\right)$

$\therefore 520 - 1.28 \left(\frac{100}{\sqrt{25}}\right)$

$\therefore 520 - 1.28 \times \frac{100}{5}$

$\therefore \underline{494.4}$

Higher fence

$\therefore \bar{x} \pm z_{\alpha/2} \left(\frac{\sigma}{\sqrt{n}}\right)$

$\therefore 520 + 1.28 \left(\frac{100}{\sqrt{25}}\right)$

$\therefore 520 + 1.28 \left(\frac{100}{5}\right)$

$\therefore \underline{545.6}$

Confidence Interval for population mean (494.6 To 545.6)

Day 6 :-

Important Topics (Inferential Stats)

Sample Data
⇓ (assumption)
population Data

* Agenda :-

(i) Z test
(ii) T test
(iii) Z test propotion population
(iv) Chi Square test
(v) Annova (F test) (Analysis of Variance)

Z-Test

Q. A factory has machine the fills 80ml of Baby medicines in a bottle. an employee belives the ~~amount~~ average amount of daily medicines is not 80 ml. Using 40 Samples he measures the average amount dispered by the machine to be is 78 ml, with a standard deviation of 2.5

~~a~~] State null & Alternate Hypothesis
b] At 95% C.I, is there enough evidence to support whether machine is working properly or not.

⇒ $\mu = 80$ ml

$n = 40$

$\bar{x} = 78$

$s = 2.5$

Z test $n \geq 30$

T Test $n \leq 30$

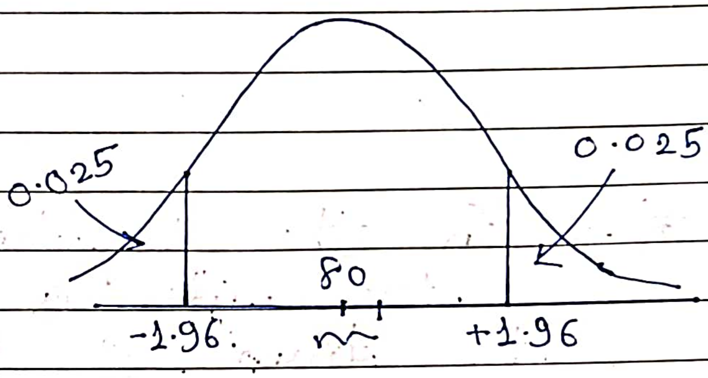
Step 1 :

$H_0 : m = 80$ (null Hypothesis)

$H_1 : m \neq 80$ (Alternate Hypothesis)

Step 2 : Significance value

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



$\therefore 1 - 0.025$

$\therefore 0.975$ (Z Table)

Step 3 : Decision making (Boundary)

Understand whether it is One tail or 2 tail test.

Step 4 : Calculates Test statistics.

$$Z = \frac{\bar{x} - m}{S/\sqrt{n}}$$

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

$$\therefore \frac{-2 \times 6.32}{2.5}$$

$$\therefore -5.05$$

Page No.
Date

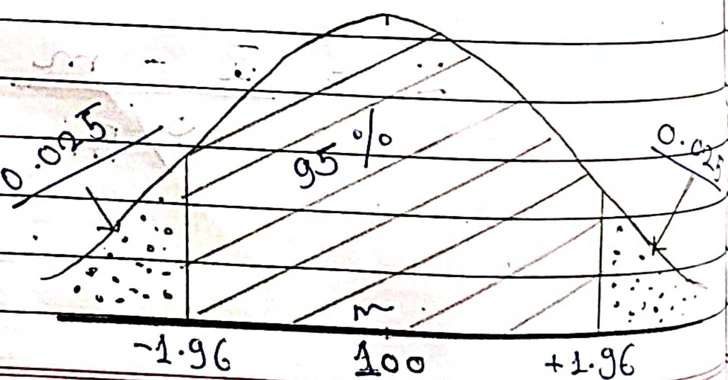
Conclusion:

If $Z = -5.05$ is less than -1.96 or greater than $+1.96$ then reject the null hypothesis with 95 C.I.

Reject Null Hypothesis {machine not working properly}

In the population the average IQ is 100, with a standard deviation of 15. A team of scientists wants 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 a medication has a mean of 140. Did the medication affect intelligence? C.I is 95%.

$\sigma = 15$
$\bar{x} = 140$
$n = 30$
$\mu = 100$



S1 : H0
H1
S2 : L
S3 :
O
S4 : Ca
Z
Z =
S5 : Co
Z

Page No.
Date

S1: $H_0 = m = 100$
$H_1 = m \neq 100$

S2: $\alpha = 0.05$ $\quad C.I = 95\%$

$\therefore \alpha = 1 - 0.95$
$= \underline{0.05}$

S3:
$\frac{0.05}{2} = \underline{0.025}$ (equally divided)

$1 - 0.025 = \underline{0.975}$ (Z table) $\underline{1.96}$

S4: Calculate Test Stats

$$Z = \frac{\bar{x} - m}{S/\sqrt{n}}$$

$$\therefore Z = \frac{140 - 100}{15/\sqrt{30}} \qquad Z = \frac{40 \times 5.47}{15}$$

$$Z = \underline{14.58}$$

S5: Conclusion

$Z = 14.58 > 1.96$

$\therefore$ The medication has done +ve effect.

Q. A Complain was register, the boys in a Government School are underfed, average weight of boy of 10 age 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 C.I = 95%. Check whether it is true or false.

$n \le 30$

Population S.D

→ $m = 32$

$S = 9$

$n = 25$

$\bar{x} = 29.5$

$CI = 95\%$

Z test Because Population S.D.

$H_0 = m = 32$ (null)

$H_1 = m \neq 32$. (Alternate)

$\therefore \alpha = 0.05$

$C.I = 95\%$

$\therefore \frac{0.05}{2} = 0.025$

$= 1 - 0.025$

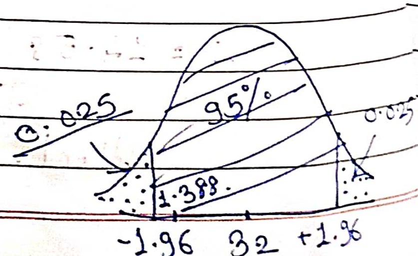
$= 0.975$

Calculate Test Statistics

$$Z = \frac{\bar{x} - m}{S/\sqrt{n}} \quad \therefore \frac{29.5 - 32}{9/\sqrt{25}}$$

$$\therefore \frac{-2.5 \times 5}{9}$$

$$\therefore -1.388$$



Conclusion ∵

∴ $-1.388 > -1.96$ So we accept the null hypothesis 95% C.I.

Students are not Underfed

One Tail

Q. A factory manufactures cars with a warranty of 5 years on the engine and transmission. An engineer belives that the engine on transmission malfunction in less than 5 year. He test a sample of 40 cars and finds the average time to be 4.8 years with standard deviation of 0.50 & C.I = 98%.

(1) State the null & alternate hypothesis
(2) At a 2% significance level, is there enough evidence to support the idea that the warranty should be revised?

$H_0 = m > 5$

$H_1 = m < 5$ {Alternate Hypothesis}

$n = 40$

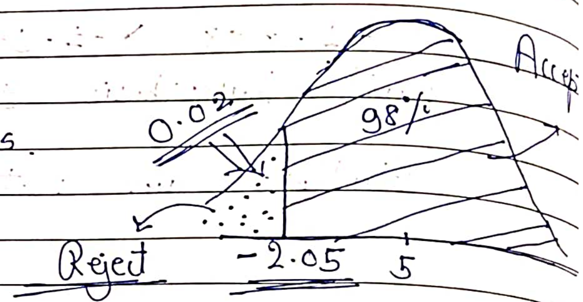
$\bar{x} = 4.8$

$s = 0.50$

(If warranty is revised less then 5 there will One tail test)

$\alpha = 0.02$

$C.I = 98\%$



Calculate Test Statistics

$\therefore Z = \dfrac{\bar{x} - m}{s/\sqrt{n}}$

$Z = \dfrac{4.8 - 5}{0.50/\sqrt{40}}$

$Z = \dfrac{-0.2 \times 6.32}{0.50}$

$Z = -2.528$

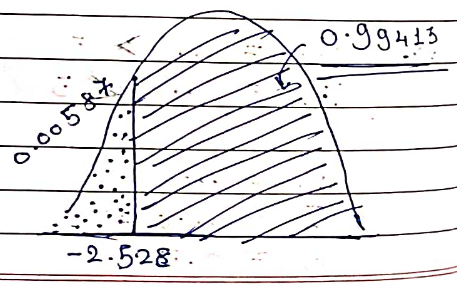
$\therefore -2.52 < -2.05$

Reject null hyptohesis

Conclusion :- warrenty needs to be revisied

* P - Value

$Z = -2.5282$



P - value = 0.00587.

$\boxed{\text{P - value} < \alpha}$

we Reject the null hypothesis

Q. The average weight of all residents in a town X, Y, Z 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 S.D 3.6

a) null & alternate hypothesis

b) 95% C.I. State whether there is enough evidence to discard the null hypothesis.

⇒ $H_0 = m = 168$

$H_1 = m \neq 168$

$\bar{x} = 169.5$

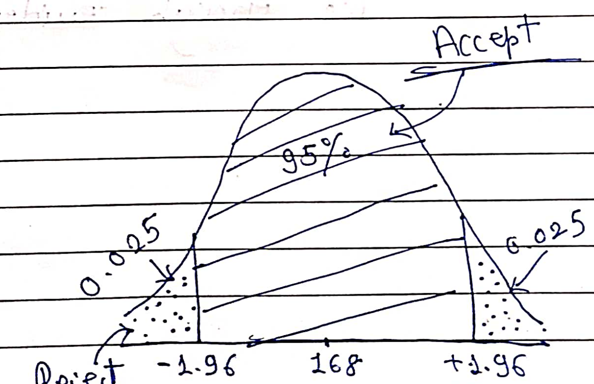
$n = 36$

$SD = 3.6$

$C.I = 95\%$

$\alpha = 1 - 0.95$

$= 0.05$



$1 - 0.025$

$= 0.975$

Calculate Test Statis.

$$Z = \frac{\bar{x} - m}{s/\sqrt{n}}$$

$$Z = \frac{169.5 - 168}{3.6/\sqrt{36}}$$

$$Z = \frac{1.5 \times 6}{3.6}$$

$$Z = \underline{\underline{2.5}}$$

Conclusion:

$2.5 > 1.96$

Reject the null hypothesis

No enough evidence to discard

Z - Test with propotion

A tech Company belives that the percentage of residents in town XYZ. that owns a cell phones is 70%. A marketing manager belives that this value to be different. he Conducts a survey of 200 individuals and found that 130 responded yes to owning a cell phone.

a) State the null and alternate Hypothesis
b) At a 95% C.I, is there enough evidence to reject the null Hypothesis

$n = 200$
$\bar{x} = 130$

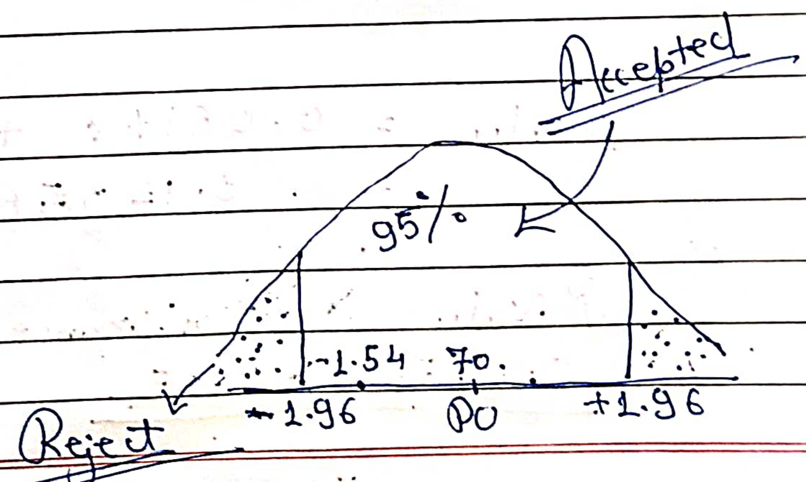
Null Hypothesis :

$$\left.\begin{array}{l} H_0 : P_0 = 70\% \\ H_1 : P_0 \neq 70\% \end{array}\right\} \quad \hat{P} = \frac{x}{n} = \frac{130}{200}$$

$$= 0.65$$

$P_0 \qquad q_0 = 1 - P_0 = 1 - 0.70$
$= 0.30$

$\alpha = 0.05$
$C.I = 95\%$



Formula for Proportion

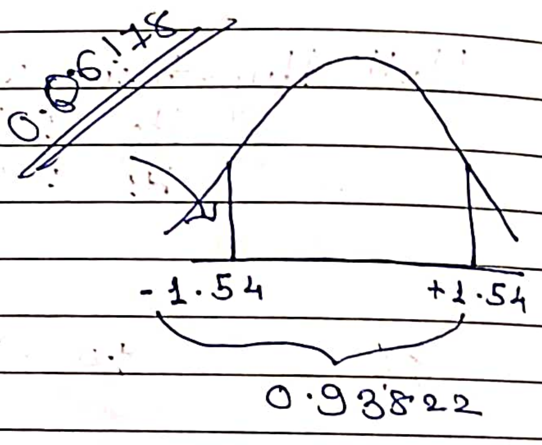
Z-test with propation.

$$Z\ \text{test} = \frac{\hat{P} - P_0}{\sqrt{\frac{P_0 q_0}{\sqrt{n}}}}$$

$$= \frac{0.65 - 0.70}{\sqrt{\frac{0.70 \times 0.30}{200}}}$$

$$= \underline{-1.54}$$

P-value.



$\therefore\ Z = \underline{-1.54}$

$\therefore\ 1 - 0.93822$

$\therefore\ \underline{0.06178}$

$$P\text{value} = 0.06178 + 0.06178$$
$$= 0.12356.$$

$\therefore$

P value > Significance value → Accept null hypothesis

$\underline{0.12356 > 0.05}$

F test

Q. A Company manufactures bikes batteries with an average life span of 2 years or more years. An engineer believes this value is to be less. Using 10 samples, he measures the average life span to be 1.8 years with S.D of 0.15.

a) State the Null & Alternate Hypothesis

b) At a 99% C.I; Is there enough evidence to discard the H_0?

⇒ $H_0 = m > 2$

$H_1 = m < 2$

$n = 10$

$\bar{x} = 1.8$

$S.D = 0.15$

$C.I = 99\%$

$\alpha = 0.01$

$\alpha = 0.01$

$= 1 - 0.99$

$= \underline{\underline{0.01}}$

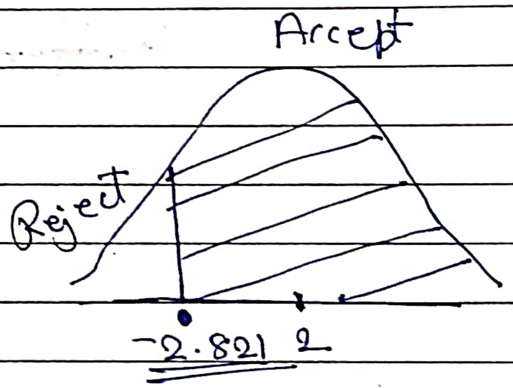
$0.01\,(9) = \underline{\underline{-2.821}}$

Degree of freedom

$\therefore n - 1$

$\therefore 10 - 1$

$= \underline{\underline{9}}$



T - Test Stats.

$$T = \frac{\bar{x} - m}{s/\sqrt{n}}$$

$$= \frac{1.8 - 2}{0.15/\sqrt{10}}$$

$$= -4.216$$

$\therefore$ $-4.216 < 2.82$ (Reject Null hypo)

$\therefore$ Conclusion.

No enough evidence to discard the Ho.

Battery life span is not for 2 yea

Assignment.

Q. A Car Company belives that the percentage of residents in a City ABC that owns a Vechile is 60% or less. A Sales manager disagrees with this he Conduct a hypothesis testing Surveying 250 residents and found that 170 responded yes to owning a vechile

a] State the null & alternate Hypo.

b] At 10% significance level, is there enough evidence to support the idea that vechile Ownership in City ABC is 60% or less?

Null Hypo : $H_0 : P_0 \geq 60\%$

$H_1 : P_0 < 60\%$

$P_0 = 0.60$

$n = 250$

$\bar{x} = 170$

$C.I = 90\%$

$\hat{P} = 0.68$

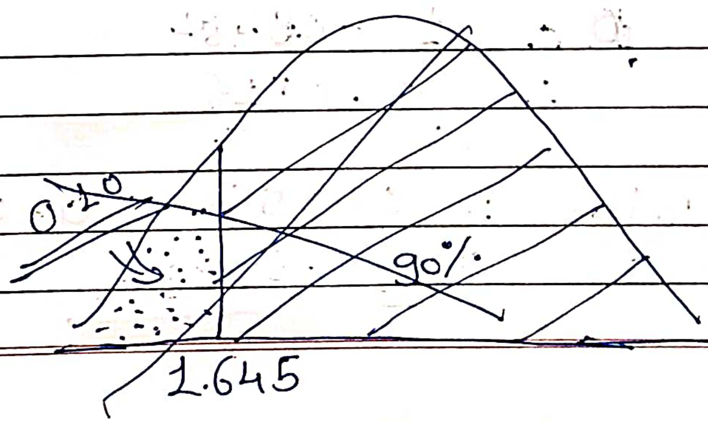
$q_0 = 0.40$

$$\hat{P} = \frac{\bar{x}}{n} = \frac{170}{250} = \underline{0.68}$$

$$q_0 = 1 - P_0 = 1 - 0.60 = \underline{0.40}$$

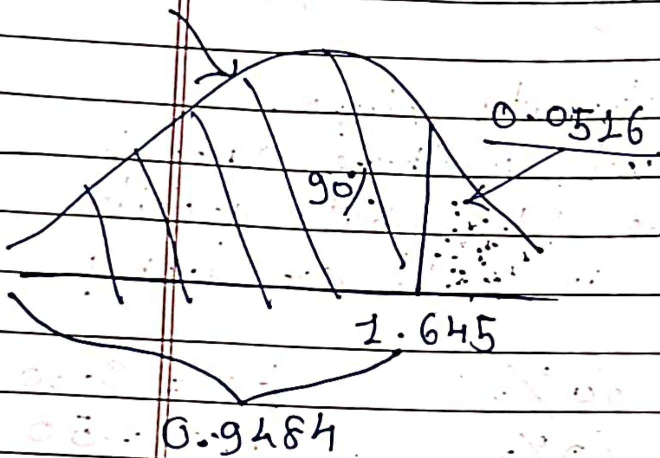
$\alpha = \underline{0.10}$

$\left\{ \because 1 - 0.90 = \underline{0.10} \right\}$



* Z Test with propotion

$$Z \text{ test} = \frac{\hat{P} - P_0}{\sqrt{\frac{P_0 q_0}{\sqrt{n}}}}$$



$$= \frac{0.68 - 0.60}{\sqrt{\frac{0.60 \times 0.40}{250}}}$$

$$= \frac{0.08}{0.04899}$$

$$= 1.6365$$

$1.6365 < 1.645$ (Accept the null hypoth

P-value : $Z = 1.6365$

$Q = 1 - 0.9484$

$= 0.0516$

P value $= 0.0516 + 0.0516$

$= 0.1032$

Conclusion :-

A residents who do not own a vechile assuming the claim made by the car company is true.