MON TUE WED THR FRI SAT SUN

Subject: Data Science

17th September, 2022

Statistics Note (Day-5)

by Krish Naik Sir

~~Agendas :-~~

Inferential Statistics :-

Agenda :-

1) Hypothesis Testing
2) p-value
3) confidence Interval
4) Significance value

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

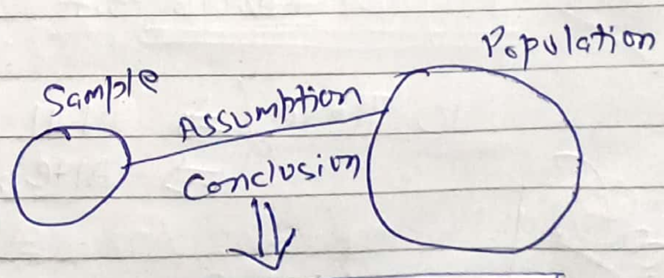
3 Distributions :-

1. Bernoullis
2. Binomial
3. Power Law

Tranformation

1) ~~Hypoth~~ Inferential Statistics :-

~~Step of Hypothesis Testing~~



⇓

Hypothesis Testing

↳ Collecting the Sample data and making conclusion about populati data using some experiment is Called: Hypothesis Testing

1. **Null hypothesis :-**

⇒ Null hypothesis comprise of default one :-

Eg. person is not a Criminal.

Step of Hypothesis Testing:-

① **Null Hypothesis** ~~Continuous~~ :-

⇓

Eg:- Coin is fair ⟶ Experiment [Coin is fair or Not]

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

② **Alternate Hypothesis :-**

Eg:- Coin is Not fair,,

③ **Perform Experiments :-**

100 times

⇓

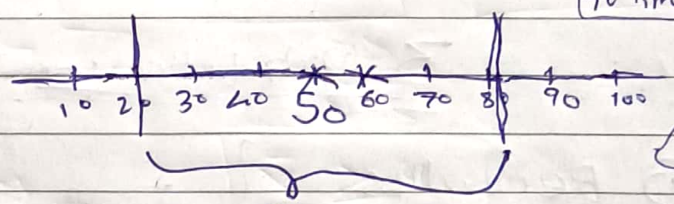
fair (50 Time Head)

~~fair~~ (60 Time Head)

70 Times ⇒ Domain Expert

⇓

confidence Interval



C.I ⇒ Confidence Interval

Eg: C.I = [20-80] ⇒ Range coming between 20 to 80 the Coin is fair

⇓

Coin is Fair

10 Time ⇒ Null Hypothesis is Rejected
⇒ Alternate Hypothesis is Accepted

Conclusion

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

→ We Reject the Null Hypothesis [outside C.I]

Conclusion

Eg:- 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, footage.

⇓

Judge

⇙

Conclusion

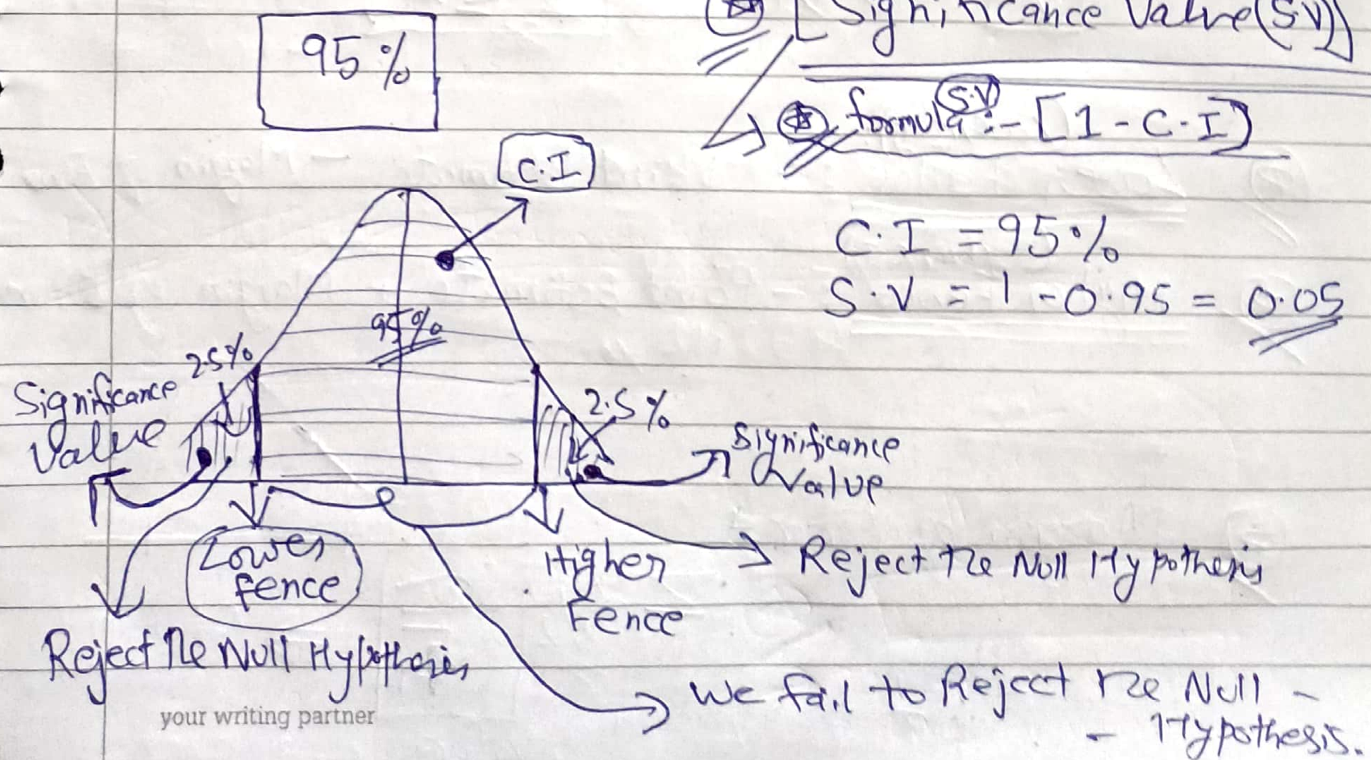
⊛ Confidence Internal (C.I):-

⊛ [Significance Value(S.V)]

⊛ formula (S.V) :- [1 - C.I]

$C.I = 95\%$

$S.V = 1 - 0.95 = \underline{0.05}$



Subject ______ MON TUE WED THR FRI SAT SUN

V.V.I. ✪ **Point Estimate** :– The Value of any Statistics that estimate the value of a parameter is called Point Estimate.

Statistics → $\bar{X}$ → Sample mean

$\boxed{\mu}$ ← Parameter → Population Means

$\bar{X} \longrightarrow \mu$

$\bar{X}$ ↓ Statistics, μ ↓ Parameter

✪ **Parameter** ⇒ population means

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

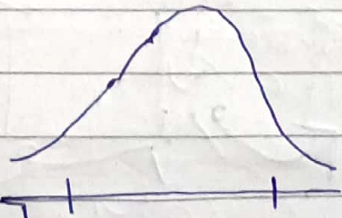
Point Estimate

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

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

✪ **Lower Fence** (Lower C.I) :– Point Estimate – Margin of Error

✪ **Higher Fence** (Higher C.I) :– Point Estimate + Margin of Error



✪ Margin of Error ⇒ $Z_{\alpha/2} \boxed{\dfrac{\sigma}{\sqrt{n}}}$ ⇒ Standard Error

your writing partner

∴ α = Significance value

Formula :-

↳ Margin of Error $\Rightarrow Z_{\alpha/2} \boxed{\frac{\sigma}{\sqrt{n}}} \Rightarrow$ Standard Error

$\therefore \alpha$ = Significance Value.

— × — × — × —

Q. On the Quant test of CAT Exam, a sample of 25 test taker has a mean of 520 with a population Standard deviation of 100. Construct a 95% C.I about the mean?

Ans Given

$n = 25$ (Sample)

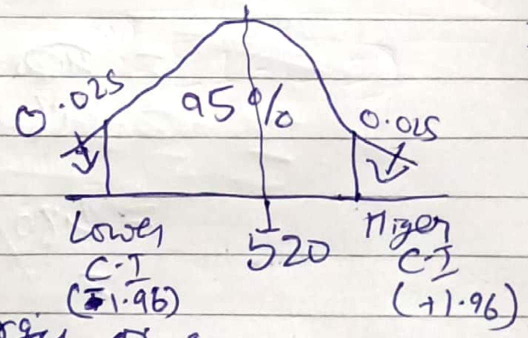
$\bar{x} = 520$ (sample mean)

$\sigma = 100$ (population Standard Deviation)

$C.I = 95\%$ (Confidence Interval)

Now, $S.V = 1 - C.I = 1 - 0.95 \Rightarrow 0.005$ //

$\alpha = 0.05$



Fail to Reject Null Hypothesis

Lower C.I = Point Estimate − Margin of Error

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

$= 520 - Z_{0.025} \frac{\cancel{100}\,20}{\cancel{5}_1}$

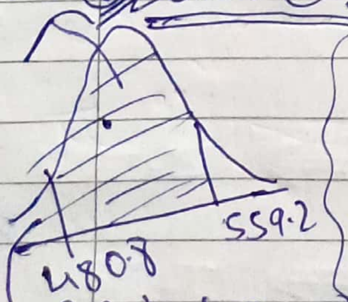
$= 520 - 1.96 \times 20 \Rightarrow 480.8$

Z-table

$\Rightarrow 0.025$

$\Rightarrow 1 - 0.025$

$\Rightarrow 0.975$



Reject the Null Hypothesis

$\therefore$ Higher C.I $\rightarrow 520 + 1.96 \times 20 \Rightarrow 559.2$ //

Q.) **Given**

$\bar{x} = 480$ (sample mean)

$\sigma = 85$ (Standard Deviation)

$n = 25$ (Sample)

C.I = 90% (confidence Interval)

Now, S.V = ~~1-0~~ 1 − C.I

$= 1 - 0.90 = 0.10$ //

Lower C.I = Point Estimate − Margin of Error

$= 480 - Z_{0.10/2} \frac{85}{\sqrt{25}}$

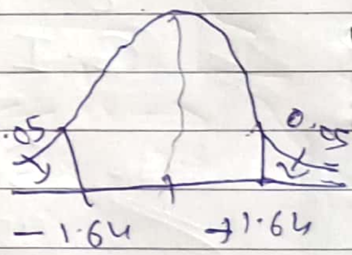
$= 480 - Z_{0.05} \left[\frac{\cancel{85}\,17}{5} \right]$

$= \cancel{480}$.

$= 480 - 1.64 \times 17$

$= 480 - 27.88$

$= 452.12$ //



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

$= 480 + 27.88$

$= 507.88$

∴ Range = [452.12 ⟷ 507.88]

Q. 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 Given

$\bar{X} = 520$

$S = 80$

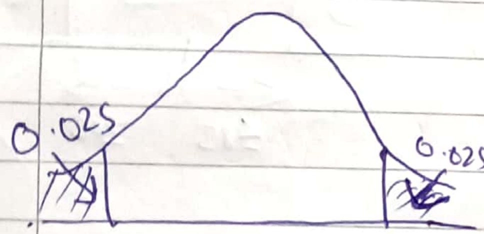
$C.I = 95\%$ $\boxed{n = 25}$

Now:

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

Now

$$\bar{X} \pm t_{\frac{\alpha}{2}}\left(\frac{S}{\sqrt{n}}\right)$$

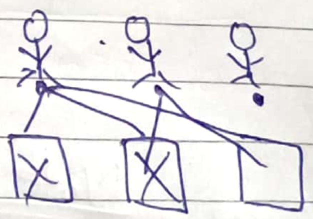


$\boxed{\text{T-test}}$

Degree of freedom

$\Rightarrow \boxed{n-1}$

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



$$\text{Lower C.I} = 520 - t_{0.05/2}\left(\frac{\cancel{80}^{16}}{\cancel{5}_1}\right)$$

$$= 520 - t_{0.025} \times 16$$

$$= 520 - 2.064 * 16$$

Lower C.I = 486.976

Higher C.I $\Rightarrow 520 + 2.064 * 16$

$\Rightarrow 553.024$

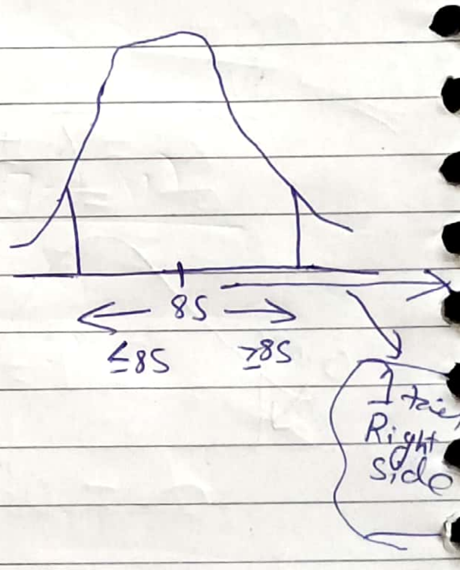
⊛ 1 Tails and 2 Tails Test :-

Q/ Colleges in 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 ?

(Two Tails)

greater than 85 % [1 tail Right side)
~~lesser than 85 % [1 tail left side]~~

Ans



Now

Z Test
T Test

Note :-

① $\boxed{\text{If } n \geq 30 \text{ and Population Std}}$ $\Longrightarrow$ Z-Test

② $\boxed{\text{If } n < 30 \text{ and Sample Std}}$ $\Longrightarrow$ T-Test

Subject ____________ MON TUE WED THR FRI SAT SUN

Hypothesis Testing

Q) A factor has a machine that fills 80 ml of Baby medicine in a bottle. An employee believe 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) State Null & Alternate Hypothesis.

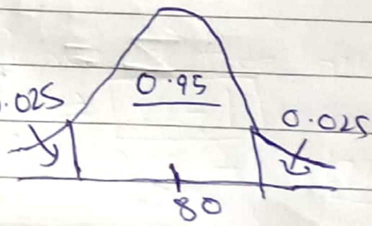
(b) At 95% C.I, is there enough evidence to support machine is working Properly or Not.

Step 1 Given $\mu = 80$ ml, $n = 40$, $\bar{x} = 78$, $S = 2.5$

Ans Null Hypothesis ; $\mu = 80$.

Alternate Hypothesis, $\mu \neq 80$

Step 2



$C.I = 0.95$

Now

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

Step 3

$n = 40$
$S = 2.5$

① [$\because n \geq 30$, or population (sd)] → Z test

② [$n < 30$. and Sample (Std)] → t test

Z-test.

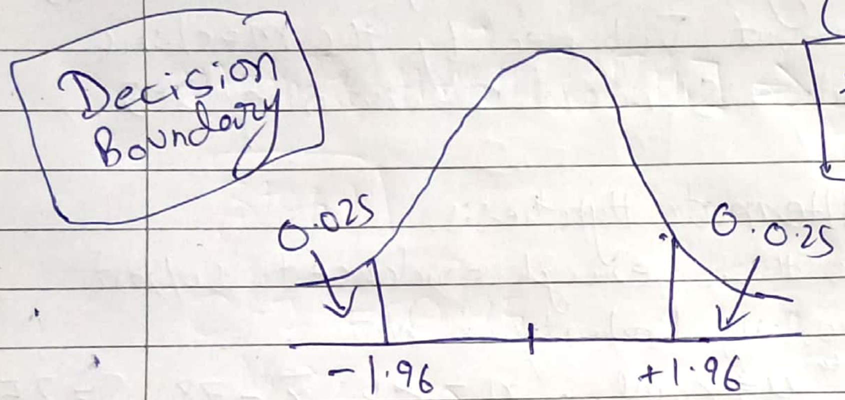
your writing partner

Step 4

Z test

Let perform the Experiment :-

$$\boxed{\begin{aligned}(S.V &= 1 - C.I)\\ &= 1 - 0.025\\ &= 0.975\end{aligned}}$$



Now
Calculate Test Statistics (Z-test)

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

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

Conclusions :-

Decision Rule :- If $Z = -5.05$ is less than -1.96 or greater than 1.96.
then we Reject the Null Hypothesis with 95% C.I

Subject

MON TUE WED THR FRI SAT SUN

Final Conclusion :-

Reject the Null Hypothesis { There is some fault in the Machine. }

Q) A Complain 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 average weight was found to be 29.5 Kgs? With C.I = 95%. Check it is True or False.

Ans) Given

· ~~A value is~~

Conditions for Z-Test:

i) We Know the population Std OR

ii) We do not Know the popution Std but our sample is large $n \geq 30$

Conditions for T-Test :-

i) We do not Know the population Std.

ii) Our Sample size is small [$n > 30$]

iii) Sample Std is given

<u>Given</u> $n = 25$, $\mu = 32$, $\sigma = 9$, $\bar{x} = 29.5$

Step 1

Null Hypothesis $[H_0] \Rightarrow \mu = 32$

Alternate Hypothesis $[H_1] \Rightarrow \mu \neq 32$

Step 2

$C.I = 0.95$

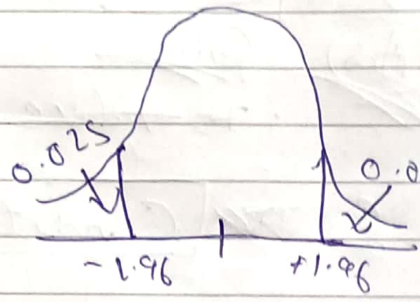
Now,

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

$= 0.05$

Step 3

Z-Test



$\underline{\text{Z-Score}} = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{29.5 - 32}{9/\sqrt{25}}$

$= -1.39$

Conclusion –

→ $-1.39 > -1.96$, Accept the Null Hypothesis

95% C.I, we fail to Reject the Null Hypothesis.

∴ The Boys are fed Wells.