(Recap) MCT – (Mean/median/mode) (Variance) (Std-dev)

Avg = $\frac{a+b+c}{3}$ = ($\bar{x}$)

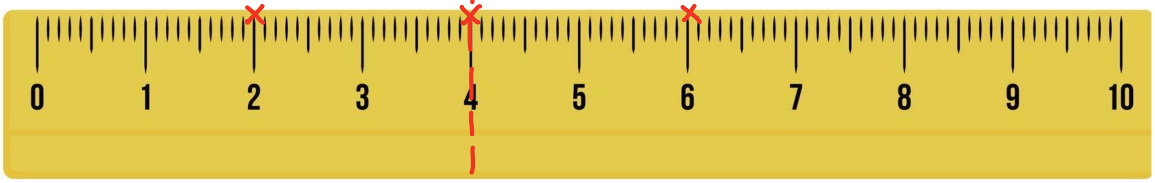
($\bar{x}$) = $\frac{2+4+6}{3}$ (Mean)

Problem

(−2)

−di

2



S → Mean

(2 −4)

(−2)

−4 ← 1 → 2 → 4

0 → 4 (2)

6 (22)

22

−6

8 (2)

−8

S/D

Spread

(Spread)

bp = v

A

(Variance) = $\boxed{\frac{\sum (x_i - \bar{x})^2}{n-1}}$

(Sample variance)

Spread

[2, 4, 8] ↑

[2, 6, 19] ↑

I love v	3000
	8000
	18,000
	2 mon

(00)

Quant

M

Vishnu

Why

but

Variance is in terms of distance squared. (Standard Deviation is a better measure to quantify the spread.)

CD

0 / Pen-d

6

4 – 2

4 :

4 — 2.2

4 — 2.2

4.4

Conn

ML

$$\frac{\sum (x_i-\bar{x})^2}{n-1}$$

$$\sqrt{\frac{\sum (x_i-\bar{x})^2}{n-1}}$$

Spread

$a_1 = 10{,}000$

$a_2 = 100$

$\boxed{a_1 > a_2}$

$\sqrt{a_1} = 100$

$\sqrt{a_2} = 10$

$\boxed{a_1 > a_2}$

$\boxed{\text{Std} = \text{Spread}}$

St — Spread ↑

EI

PCA / SVD

Joy

V St } goal

Clear

I

V

St

2002

Pop

25,000+

Sample

500

600

(Survey)

(Are you happy Campus!)

No

Valli = 500

Urry = 650

Khan = 750

An = 800

200

50

692 — good

google

Power analysis

↑ = σ

isnt it better to take maximum sample

σ =

1mt

90

10

T

S1

2000

85 — C — Repre

15 — T

C1

biased Samp

S2

55,000

50 — C

50 — T

1

Size | Sample

good

2000

5000

7000

10000

Rep

100%

1.4

google = Avg Indi = (5.9) = 1.4B

= (5.5) — 4

R R => V V VV LS

(google)

1,00,000/- — 5000 — 87,000/-

Sam Cent

Pop

S

C

Yes

Possible

(Pop) = VL | VS

1.4B

126 — Pop

Pop mean

pop std

pop mode

pop str

Pop Param

Sample $\bar{x}$, m, m, v, s

Sample Statistic

(Normal distribution)

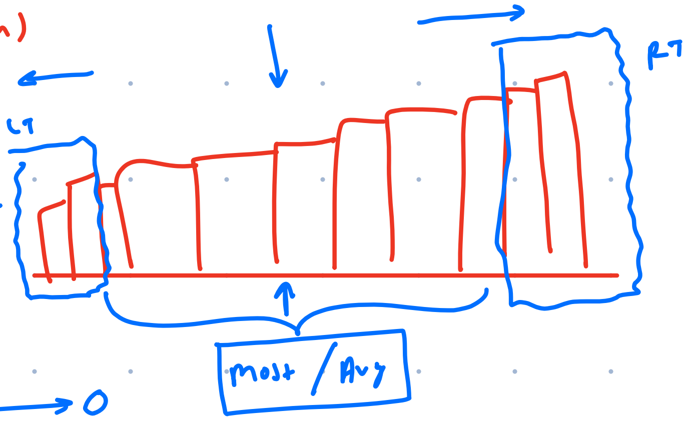
Recap

(Village)
↓
(100)

Stat

Pattern

100



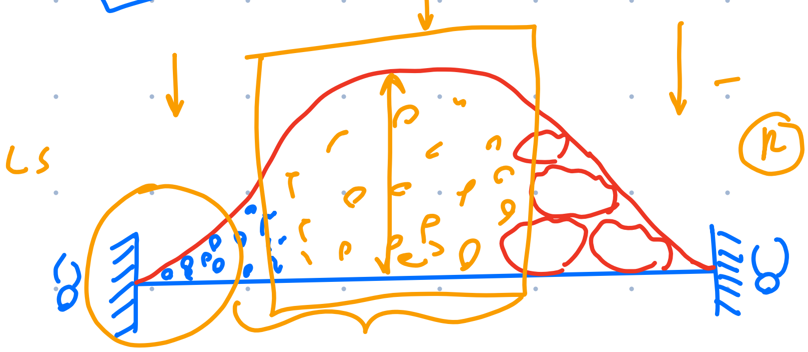
H → O

Aditya

Jay

Rock
Stone
Pebble

Density Curve



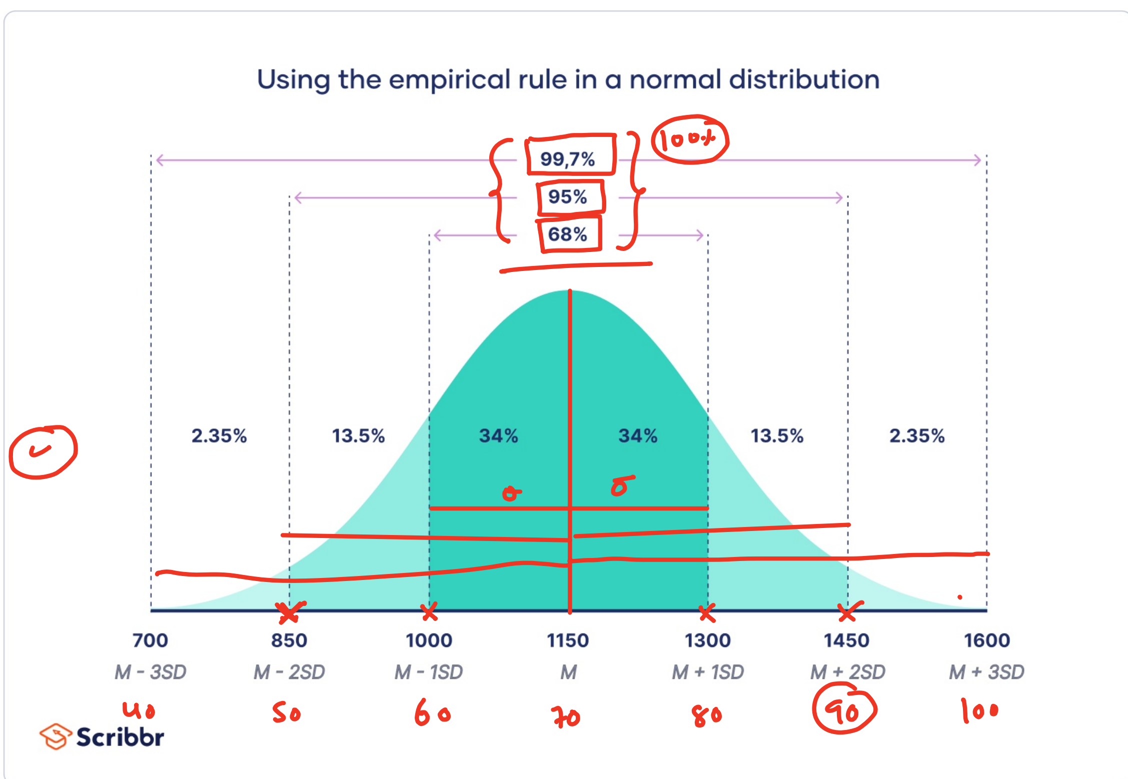
Play

Rule → DR

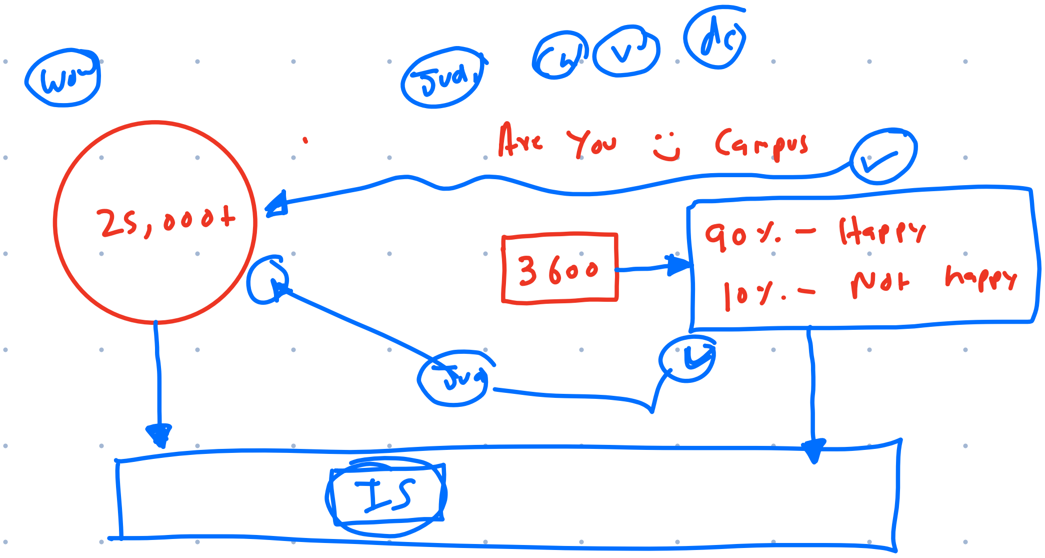
68/95/99.7 Empirical rule

100
↓
70%

Std = 10



(60% – 80%) = 68



While descriptive statistics describes the data, inferential statistics is used to draw conclusions about the population based on statistical findings on sample analysis.

Judge

Who: pop

Sample

Clah

V

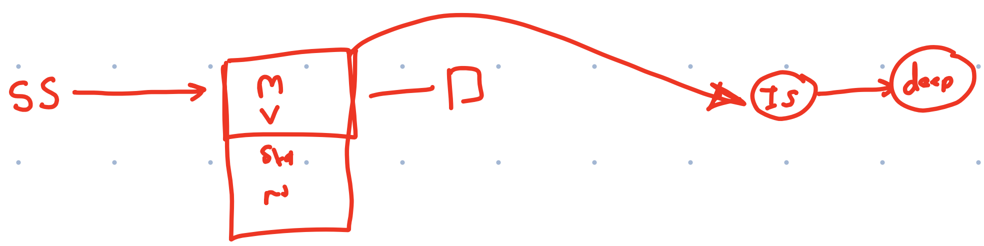
Joy

Is = 100%

No

Calc A

whats the diff bitween in inferential and sample statistics



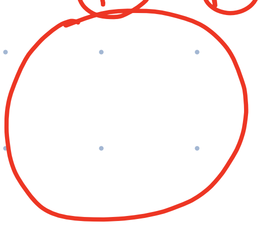
life

Covid = 19

Rel

Pun

1



=>

α PS =>

PA ⇔ SA

~~SR~~

2 Avg problem = Vill

126 = 118.167

Caten

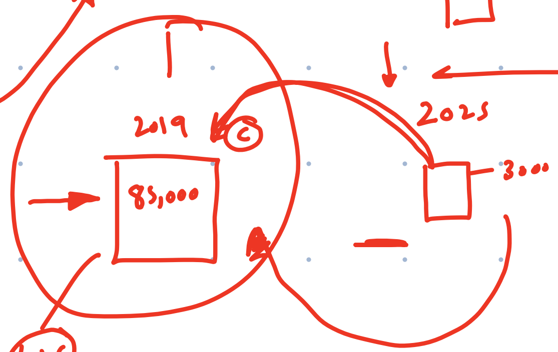
3 (1.4B) →

good

Is

~~Brand? Pro~~

RW



ER

T

b

P

P

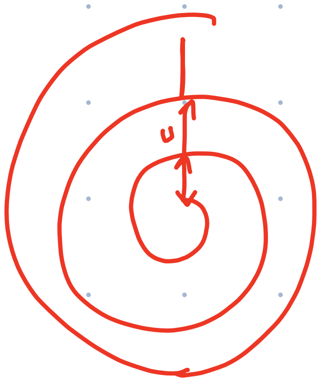
gov

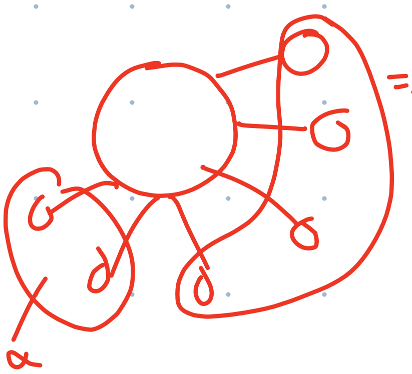
MR

N

5+

Fun





=> SS

DS = RS
PS
G.R

(Imp) = VVVVI

(learn to take a break)

1. Random
2. Purposive
3. Stratified

revise

read

here

DS ⇒ Check

Materials

(C I)

UB/LB

$$\boxed{C.I = \bar{x} \pm z \cdot \frac{s}{\sqrt{n}}}$$

$\bar{x}$ = Sample mean

z = z critical value

s = Std dev

n = Sample size

$\boxed{1.4B}$ → (Avg protein)

162.35

$\boxed{\text{Pop mean}(\mu)}$

Judge

How

Problem

1. CI is never calculated you assume

Ex: 95%, 99%, 65%, 99.25%.

(upper bound)

$$UB = \bar{x} + z \cdot \frac{s}{\sqrt{n}}$$

$$LB = \bar{x} - z \cdot \frac{s}{\sqrt{n}}$$

lower bound

122 + 14

122 − 14

Pop

PA

55 men) — on — 122 grm with std = 14 grm

Interval at 95% Confidence — 99

change

Conf

Data: $\bar{x}$ = 122 g

σ = 14 g

$n = 55$

$\boxed{z = 1.96}$ — assume

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

$UB = \bar{x} + z \cdot \frac{s}{\sqrt{n}} = 122 + 1.96\left(\frac{14}{\sqrt{55}}\right) = \boxed{125 \text{ grm}}$

$LB = \bar{x} - z \cdot \frac{s}{\sqrt{n}} = 122 - 1.96\left(\frac{14}{\sqrt{55}}\right) = \boxed{118 \text{ grm}}$

CI

(Range)

Could B

μ

125

118

$\boxed{95\%}$

Clear

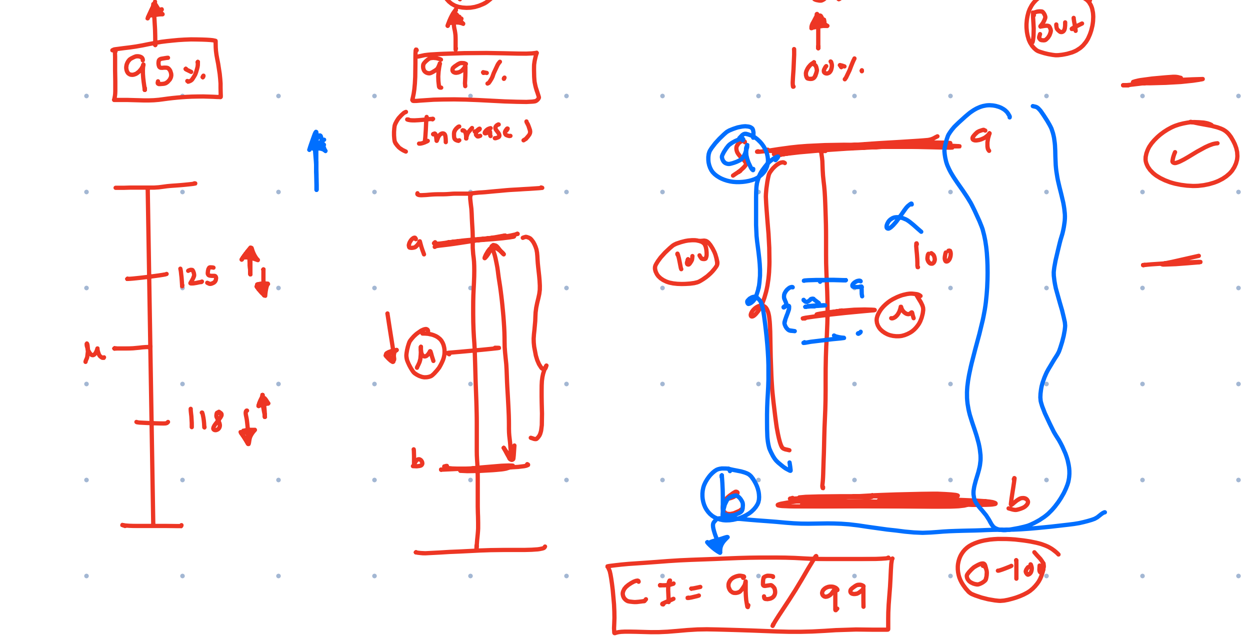
5%

1%

Sample — Pop

$\boxed{\text{CI - range of val likely to contain pp}}$

0%



(CI) = risk = No

(Hyp Test)

Komal — berlin — NY
Apoorva — NY
10

Komal → (Apoorva still same)
Komal → (Apoorva Not same)

No Change (Same)

Eve Chy (same)

(1)

Hyp

PS → Null Hyp (H_0) ⟶ Status quo — (No Change) past p F

PS → Alternate Hyp (H_1/H_a) ⟶ (EC)

(Ex)

(2020 ⟶ 85,000/-) — (Conducted) pop — Salary

2025 ⟶ 70 ⟶ 97,000/-

H_0 : (μ) = 85,000/-

H_a : (μ) $\neq$ 85,000/-

Ex (Choc) ⟶ (100 gram) Std

1,00,000/- — 100 ⟶ 99.8 gm

$H_0 : \mu = 100$

$H_a : \mu \neq 100$

(Remember)

P — Ho / Ha ⇒ t / z / f / χ^2 / A → Ho - true ; Ha - true

Ho - T	Ho - f
Ha - f	Ha - T

Prob — Ho / Ha

Stat — True / false

(Accepted / Rejected)

No — A t
— A f

Errors

(ground truth)

Komal

Kind / Soft

(AP)

Apoorva (P)

Not kind / Not soft ⇒ (error)

Nandina (P)

Kind / Soft

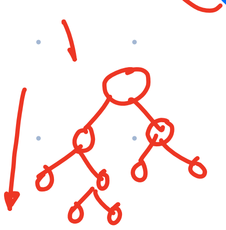
Pat: (+) Doc: (+) (True positive)	Pat: − Doc: + (False positive)
Pat: (+) Doc: (−) (False Negative)	Pat: − Doc: − (True Negative)

Type-I / α /

Type-II

Error

B — T1, T2



1000

T-II

(legal) (NG) — TI

→ (G)

Problem

R	A
Ho – T	Ho – T
Ha – T	Ha – T
Ho – T	Ha – T — Type – I
Ha – T	Ho – T — Type – II

ISC = ST

TI

EX

2021 ⟶ 150
2025 ⟶ 185

Ho : $\mu = 150$ — P

R — Ha : $\mu \neq 150$

S <= 30

(T-test)

ND

Sample Groups

T-test is a parametric test, that compares the means of the two ~~samples~~. Ideally, a sample for t-test should have less than 30 values. There are a few other assumptions that are taken before we can conduct a t-test.

Assumptions

1. The samples are independent
2. Homogeneity in sample variances
3. The Data is assumed to be normally distributed.

—— Slow ——

Re | MCT S | P/S | $CI = \bar{x} \pm Z \cdot \frac{S}{\sqrt{n}}$ UB = LB = | HT Ho Ha | EHT TI = FPE TII = FNE

PP

(PhonePe)

(Cup) (Add)

(UB – LB) 95%

$(CI - \bar{x} \pm Z \cdot \frac{S}{\sqrt{n}})$ | HT Ho Ha | TI TII

T

25,000+ — Pop

S_1 — 30 — 99.2% > S_2 — 30 — 87.17%

$\overline{S_1} > \overline{S_2}$

T = ST

1. Means given
2. Compare

$a_1 \geq a_2$

$a_2 > a_1$

$a_1 \neq a_2$

$a_1 \approx a_2$

$a \geq a_2$

$a_2 >= 1$

Significance

D	ND
1g	10g

0.01

SH

ND

Data = ST PS

PS

(RA) = Res

Pha

T-test

1. One Sample
2. Ind " T-test
3. Paired T-test

One-Sample-t-test

(Chocolate) — Bar = 100g

1,00,000+ — 30

1,00,000 — 100g, 30 — 99.7

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

$\bar{x}$ = Sample mean

μ = Pop mean

s = Sample std_dev

n = Sample size

Goal

Ind Sample T-test

P1 — P2

S1 ⟷ S2 Ind

USA ⟷ Bharat

30 — 30

99.6 — 99.99%

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Paired T-test

P

diet pill

	Before	After	
P1	90	85	a
P1	100	80	b
P1	90	70	c
P1	65	55	d

effect ENO / E S

$$t = \frac{\bar{x}_d}{\frac{s}{\sqrt{n}}}$$

(T-test) => T-distribution

1 z-test (dof)

H0

Covid-19 = R/S/1

pop

Avg

α

Do You know pop std-dev

Yes → Sample size > 30

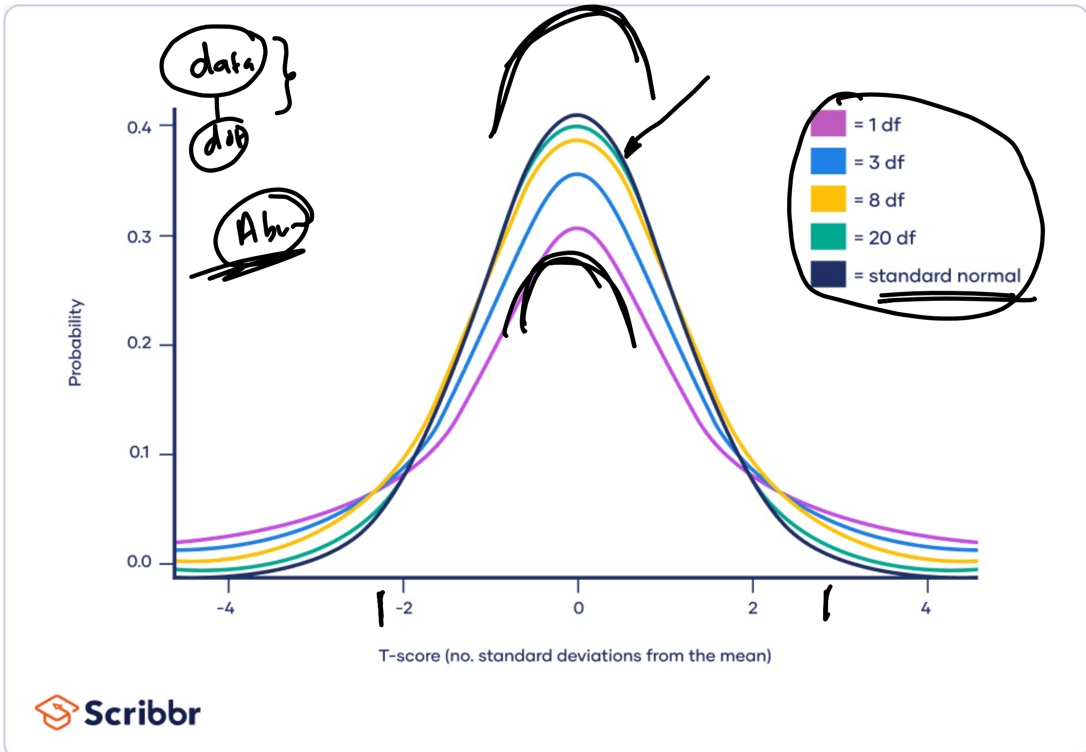
- Yes → Z-test
- NO (≤30) → T-test

NO → Use T-test → No

100

Ex

(n-1) = df



Error

gain

100%

reporak | Doll

rep RBI

< Ghana

lap pu Bullh

(Independ) : 30 30 1

H0 => □

Bree

α

(done) (13 pg) One-Sample-T

$\boxed{Pop(\mu) = 100}$

The average salary of a person is hundred dollars back when the survey was conducted in 2020 in fresh survey conducted in 2025 with 30 participants, it was found out that the average salary of a person is one $40 with a standard deviation of $20 can you calculate whether this is a significant difference in salary or not at a confidence interval of 95%?

Data: Pop-mean(μ) = 100 | $n = 30$ | CI = 95%

Sample mean($\bar{x}$) = 140 | $S = 20$

① (Hyp)

$H_0 : \mu = 100$ (No Significant Change in Salary)

$H_a : \mu \neq 100$ (Sig Changes in Salaries)

(>, <) $\boxed{\text{Two-tailed - t-test}}$ ②

② Significance level (α)

α = 100% − CI%

= 100 − 95

$\boxed{\alpha = 5\%}$ or $\boxed{\alpha = 0.05}$

③ degrees of freedom $\boxed{df = n-1}$

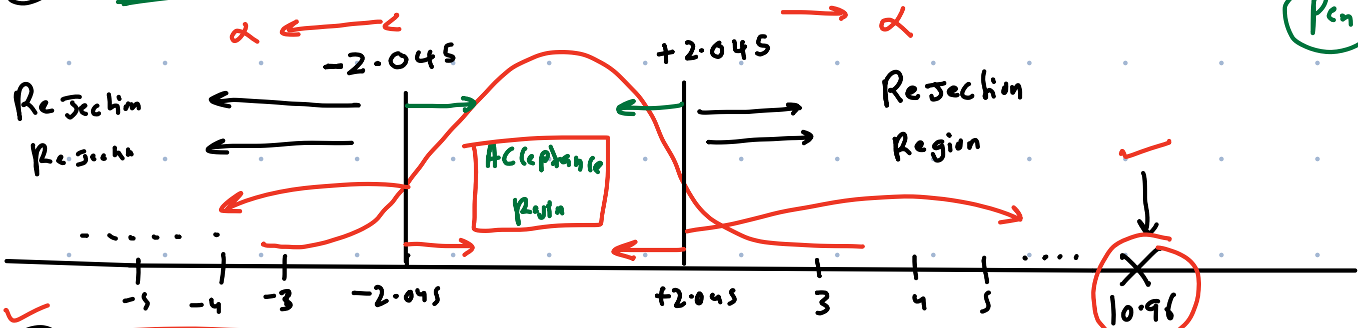
$\boxed{df = 30-1 = 29}$

④ (Boundation) (T-Critical_value) = 2.045

(t_s > t-c) − R

(t-s < t-c) − R

α Pcn



⑤ $\boxed{\text{t-Statistic}}$

$$\boxed{t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}} = \frac{140 - 100}{\frac{20}{\sqrt{30}}} = \boxed{10.95}$$

$\boxed{\text{Rej } H_0}$

NO

α CI t

cdf α ✓ α

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.99}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390

✓

Pdf
cdf
ppf

c → ppf → P P F = Auc

One tail

from Bibhu Mohanty to Rudra (privately): 11:41 AM
Sir, how do we decide this is two tailed? what does that mean and is it coming from the assuption we took(>,<) than sample mean?

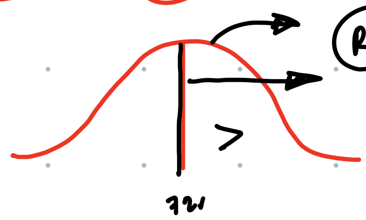
from kavita sharma to Rudra (privately): 11:41 AM
after 30 no. gap is 10 in table why?

from Kaushal to Everyone: 11:41 AM
What is source of T - Table? I sthere any formula to calculate?

Kaushal = 72%

Svv = > = Better

2 0



O/T

P eos

eq
ed } σ

Billone => 350 kg

(Puspa) = 150

LT

A

Pen/Pap

Note: $\boxed{t\_c \;/\; t\_s}$

α α

Code

p_value < alpha

✓

data

$\boxed{CI = 95\%}$

t-stat

$\alpha = 5\%$

w

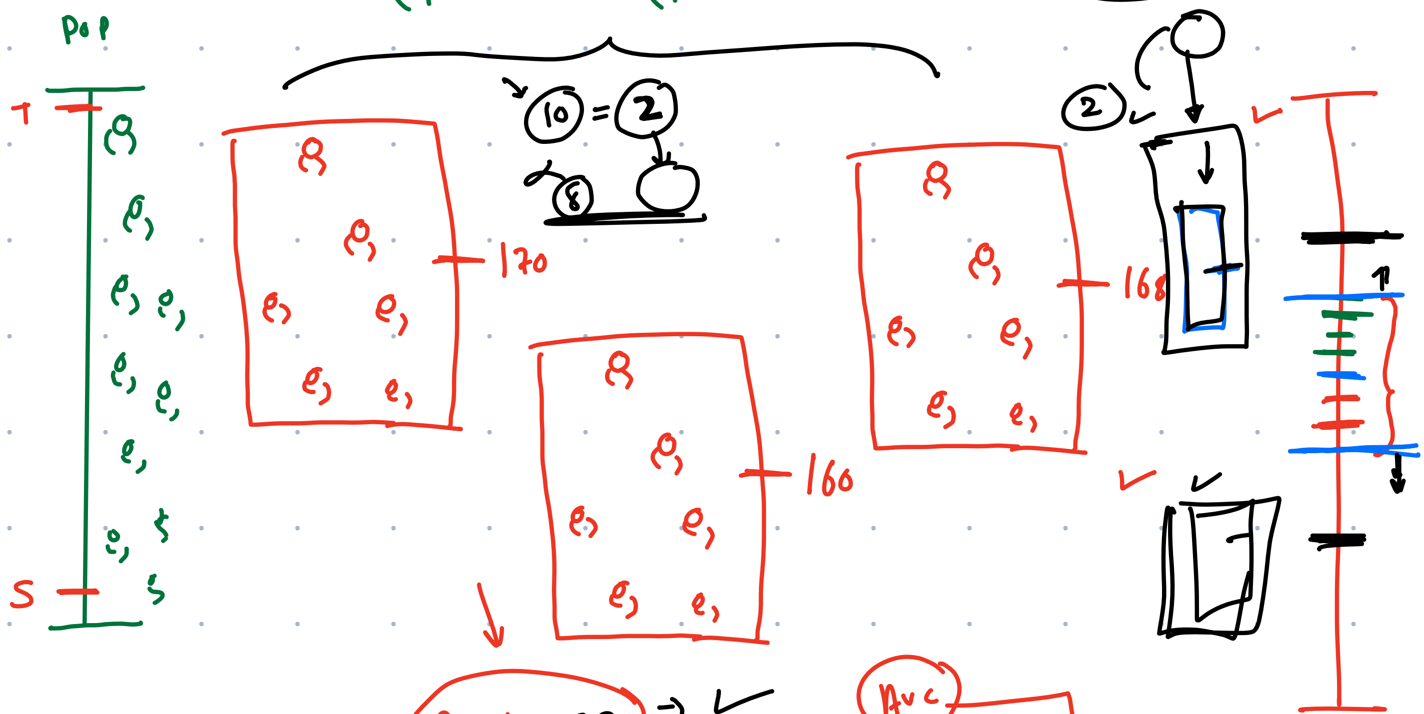
$\boxed{\text{t-critical value} = 2.045}$

✓ e/he

$\boxed{\text{p\_value} =}$

(p_value) = (Probabili valu)

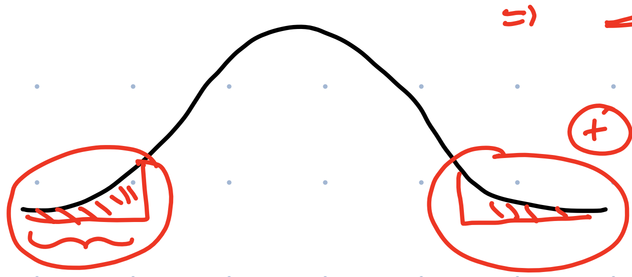
Extreme



p_value = 20.. ⇒ ✓

Avc

⇒ + = $\boxed{\text{p\_value}}$



+ ⇒

100 − 60 =

Claim

HD — 30 min

Older = 60%

Cal p ⇒ $\boxed{\text{Fischer Exact Test}}$

PPE ∝ α

$$p = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{a!b!c!d!n!}$$

$\boxed{\begin{matrix} a \\ b \\ c \\ d \\ n \end{matrix}} = CT$