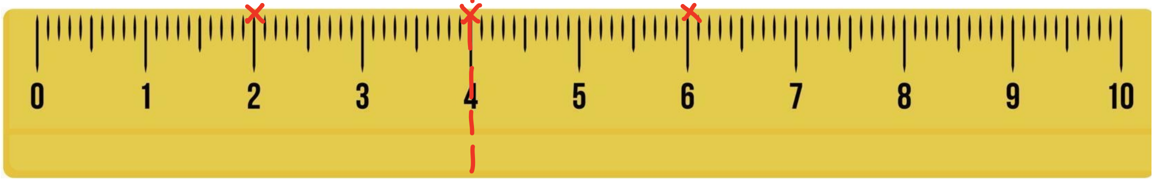
(Recap) MCT - (Mean/median/mode) (Variance) (Std-dev)

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

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

(Problem)

(-2) -d1 2



S → (Mean)

(2 - 4)
(-2)

-6

-8

-4 ← 1 → 4, 0 → 4, 0 → 6, 0 → 8 (2)

(Spread)

S/D

Spread

bp = v

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

(Sample variance)

Spread

[2, 4, 8] ↑

[2, 6, 19] ↑

I love	3000
	8000
	18,000
	2 mil

(∞)

Quant

M

Vishwas

Why

but

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

CO

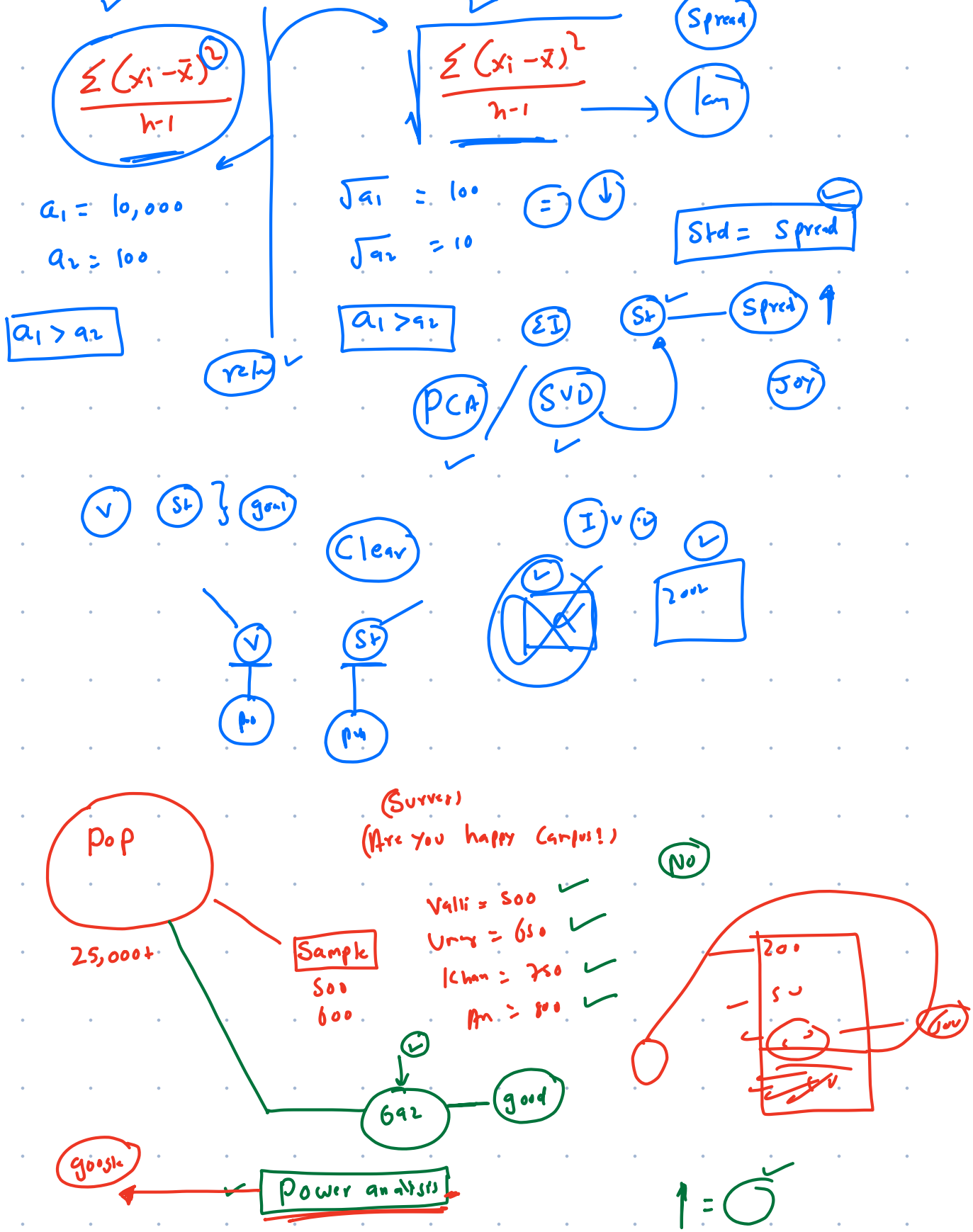
0 / Pen-d / 6

4 - 4 : 2

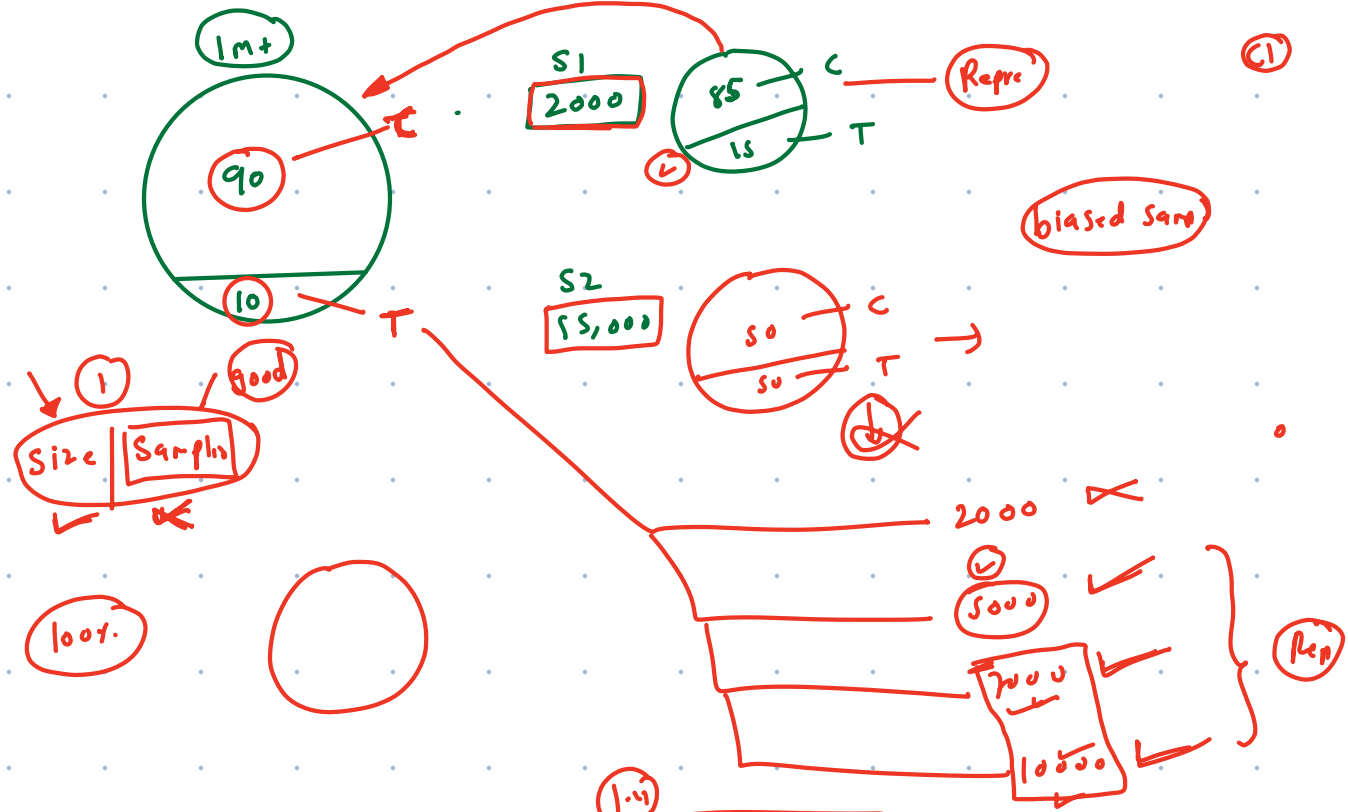
4 — 2.2
4 — 2.2
4.4

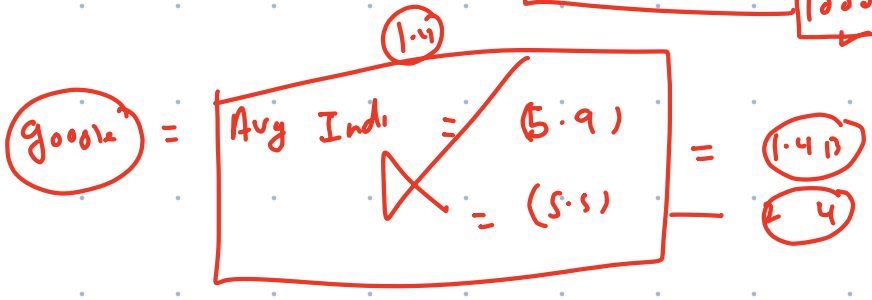
Corr

ML

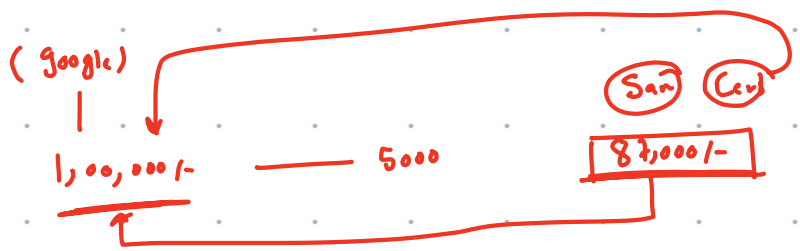


(isnt it better to take maximum sample)





R R => V V V V L S



Pop

S

C

Yes

Possible

(Pop) = V L | V S

1.4 B

126 - Pop

Pop mean

pop std

pop mode

pop str

Pop Param

Sample $\bar{x}$

n

m

V

S

Sample Statistic

(Normal distribution)

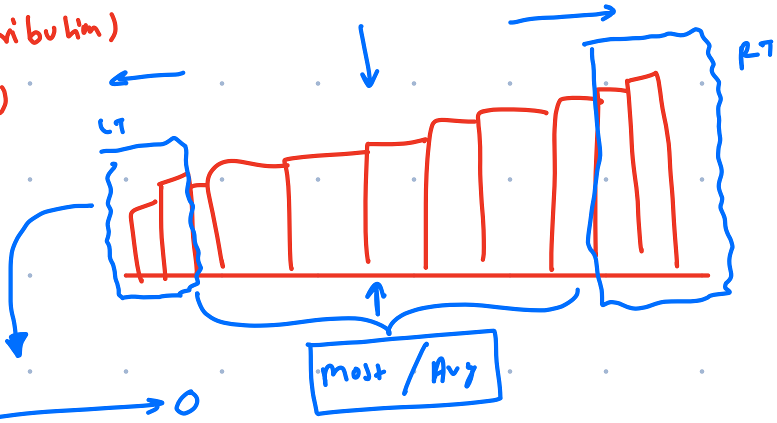
(Village)
↓
(100)

Recap

Stat

Pattern

100

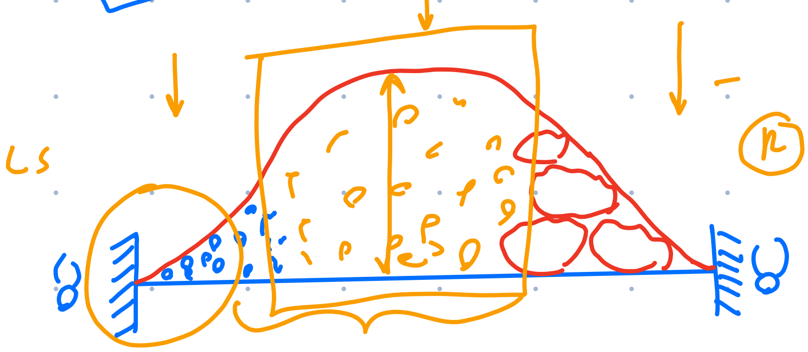


H → O

Aditya

Rock
Stone
Pebble

Density Curve



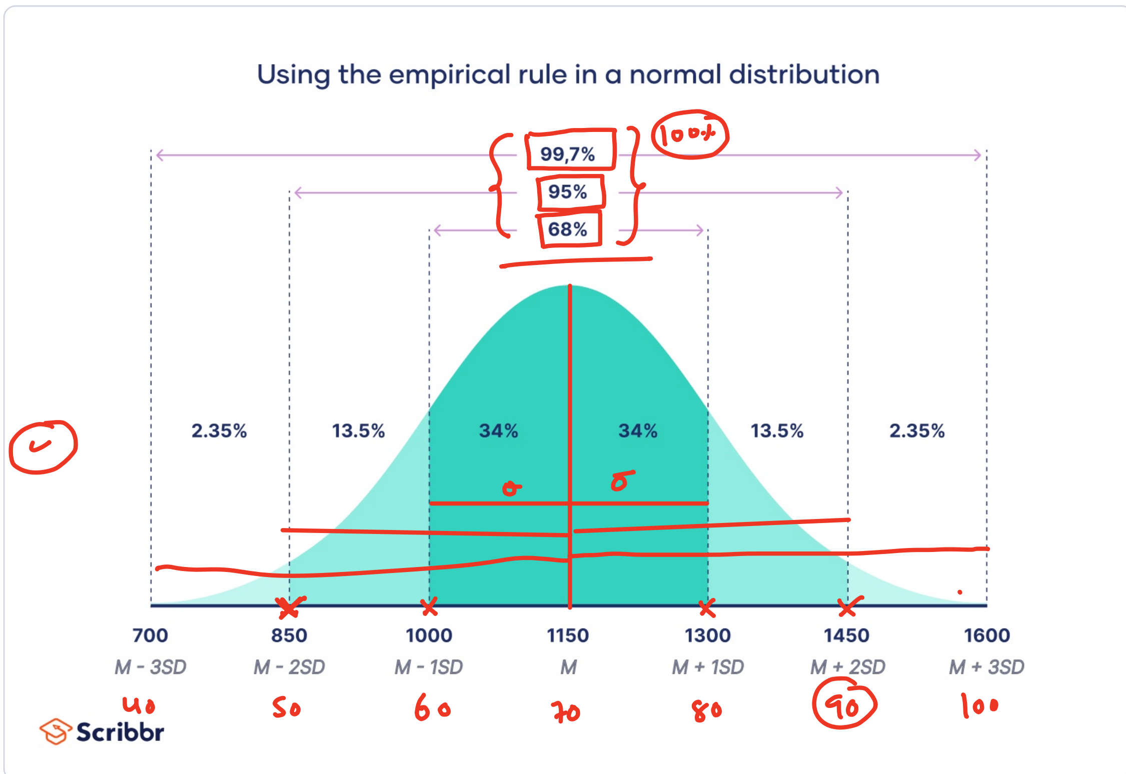
Play

Rule → DR

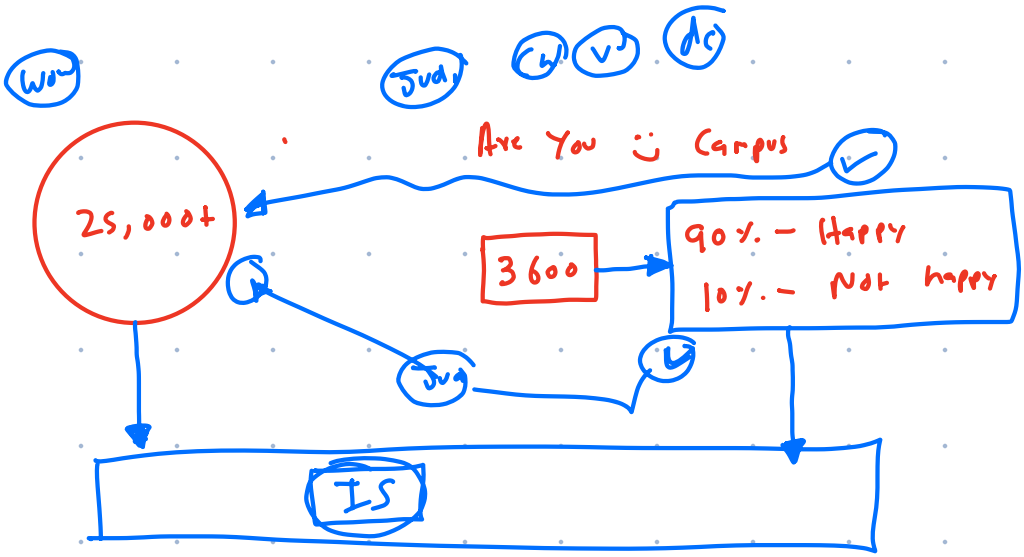
68/95/99.7 Empirical rule

100
↓
70%

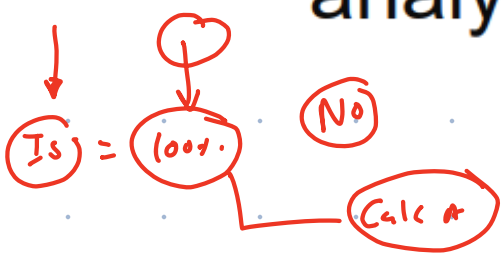
Std = 10

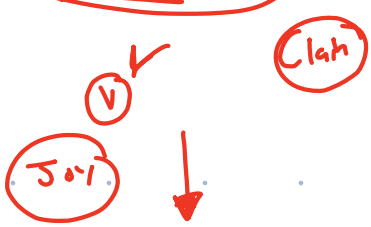


(60% – 80%) = 68

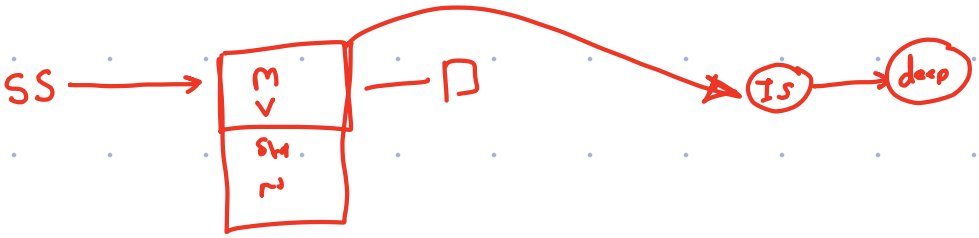


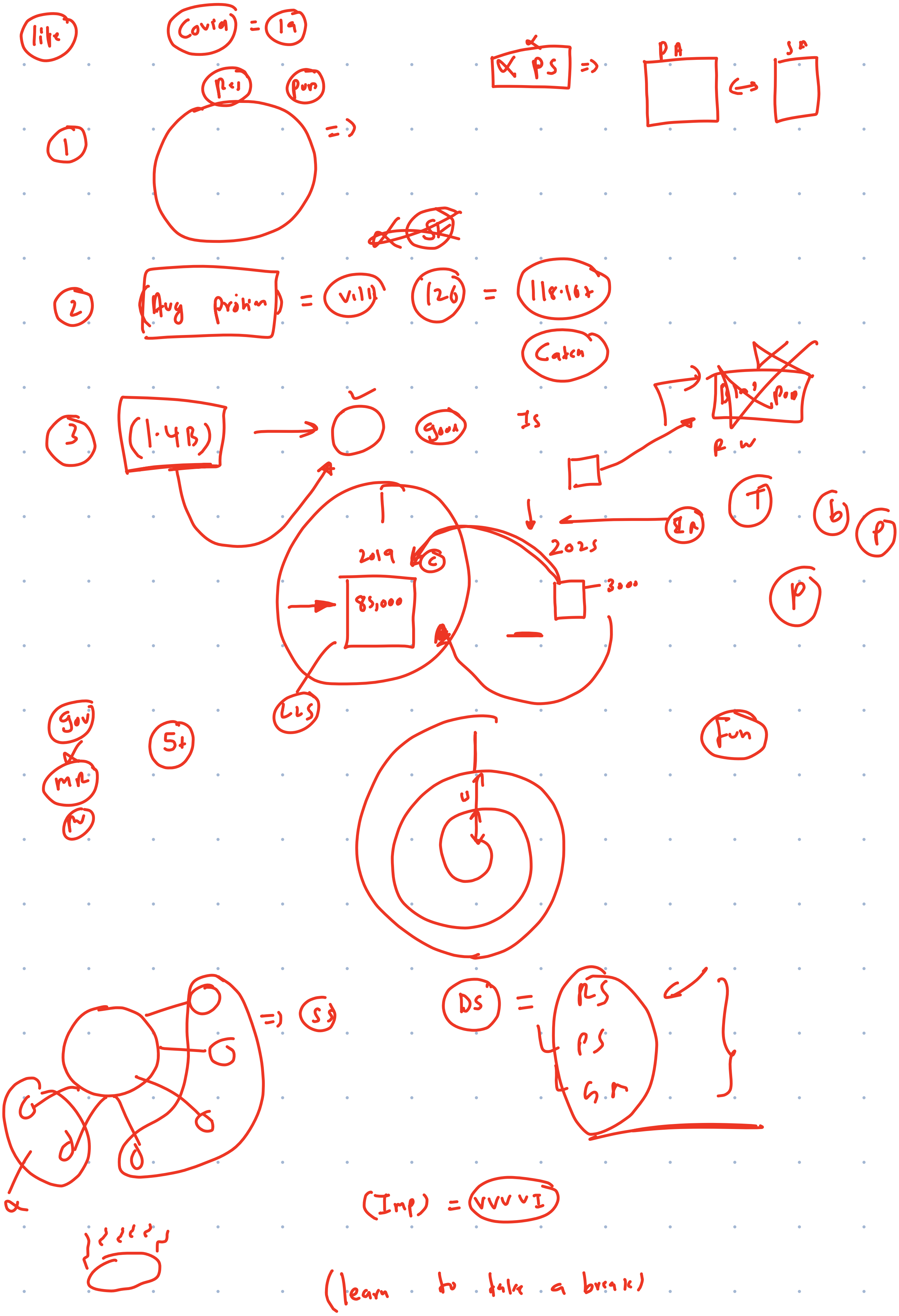
While descriptive statistics describes the data, inferential statistics is used to draw conclusions about the population based on statistical findings on sample analysis.





whats the diff bitween in inferential and sample statistics





life
Covid = 19
PS =>
PA
SA
1
2
Avg protein
126 = 118.167
Catch
3
(1.4B)
good
2019
85,000
2025
3000
gov
MR
5+
Fun
DS = RS
PS
S.R
(Imp) = VVVVI
(learn to take a break)

1. Random
2. Purposive
3. Stratified

(revisit) (read) (here) (DS) => (Check) (Materials)

(C I)

UB/LB

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

$\bar{x}$ = Sample mean

z = z critical value

s = Std dev

n = Sample size

1.4B → (Avg protein) → (162.75) → (Judge) How (Problem)

Pop mean (μ)

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)

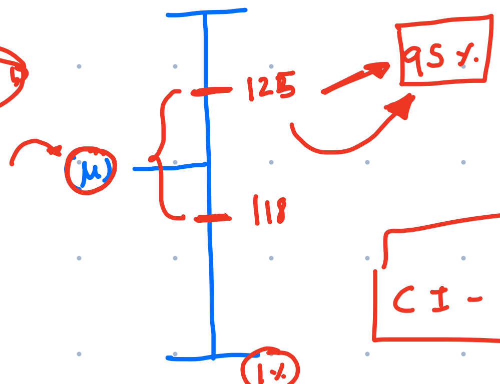
Data: $\bar{x} = 122$ g
$\sigma = 14$ g

$n = 55$
$z = 1.96$ — (assume) → Z-test

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

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

(CI)



(Could B)

(Clear)

5%

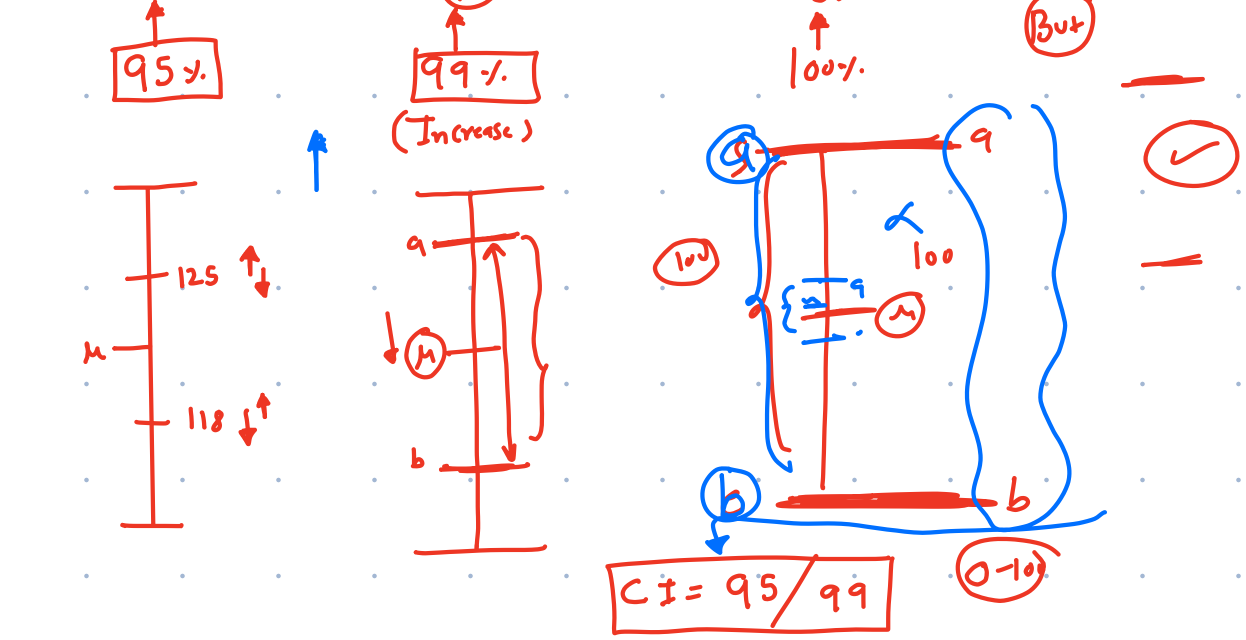
(Range)

(Sample) — (Pop)

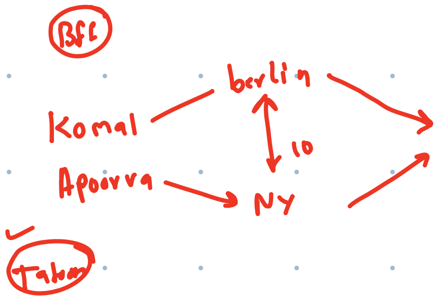
CI - range of val (likely) to Conf PP

1%

0%



(CI) = risk = Not



(Hyp Test)

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

(1)

No Change (Same)

Eve Chg (same)

Status quo – (No Change) past P F

Hyp

PS → Null Hyp (H_0)
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) ≠ 85,000/-

Ex (Choc) → (100 gram) Std

1,00,000/- — 100 → 99.8 gm

H_0 : μ = 100

H_a : μ ≠ 100

(Remember)

p → H_0 / H_a ⇒ t / z / f / χ^2 / A → H_0 - true ; H_a - true

H_0 - T	H_0 - f
H_a - f	H_a - T

Prob — H_0 / H_a

Stat — True / false

(Accepted / Rejected)

No — A + ; — A f

Errors

(ground truth)

Komal: Kind / Soft (AP)

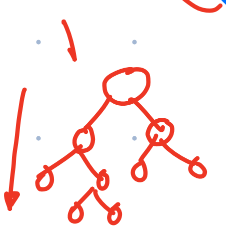
Apoorva (P): Not kind / Not soft ⇒ (error)

Nandina (P): Kind / Soft

Pat: (+) Doc: (+) (True positive)	Pat: – Doc: + (False positive) → Type-I / α /
Pat: (+) Doc: (–) (False Negative) → Type-II	Pat: – Doc: – (True Negative)

Error

B — T1 ; T2

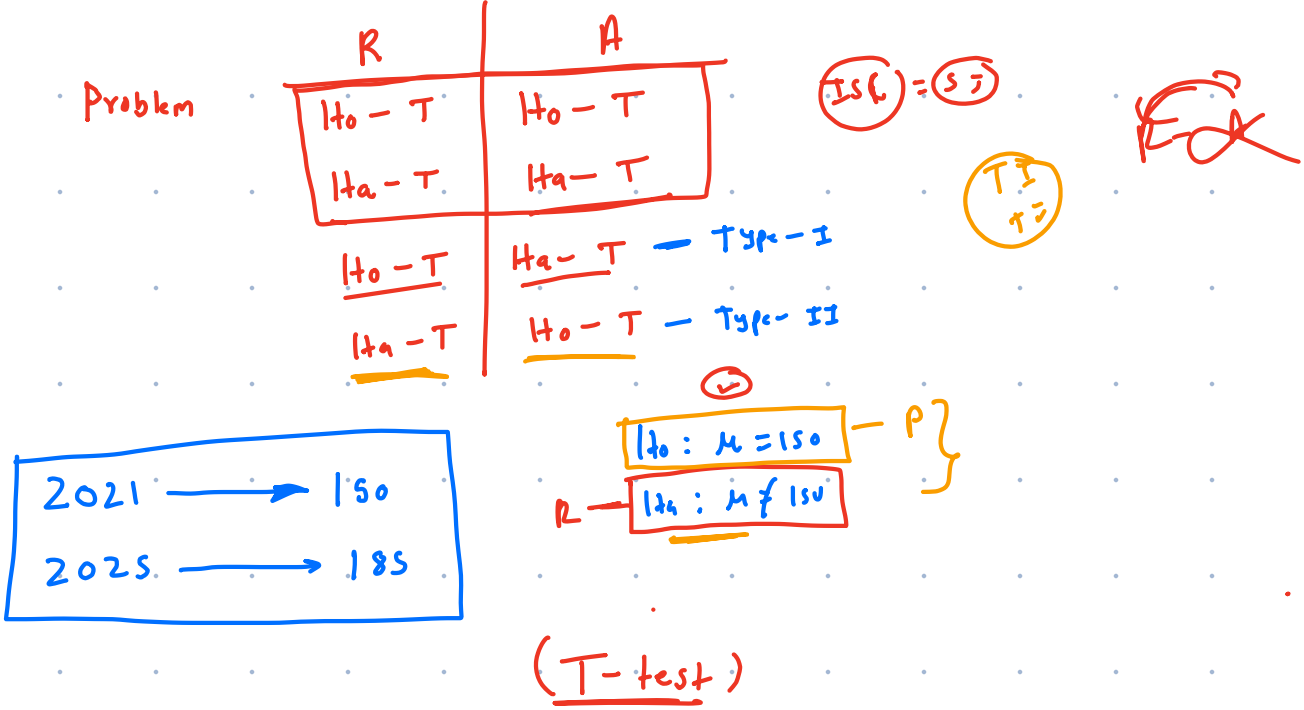


1000

T-II

(legal) (NG) — T I

→ (G)



T-test is a parametric test that compares the means of the two ~~samples~~ Sample Groups. ~~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.