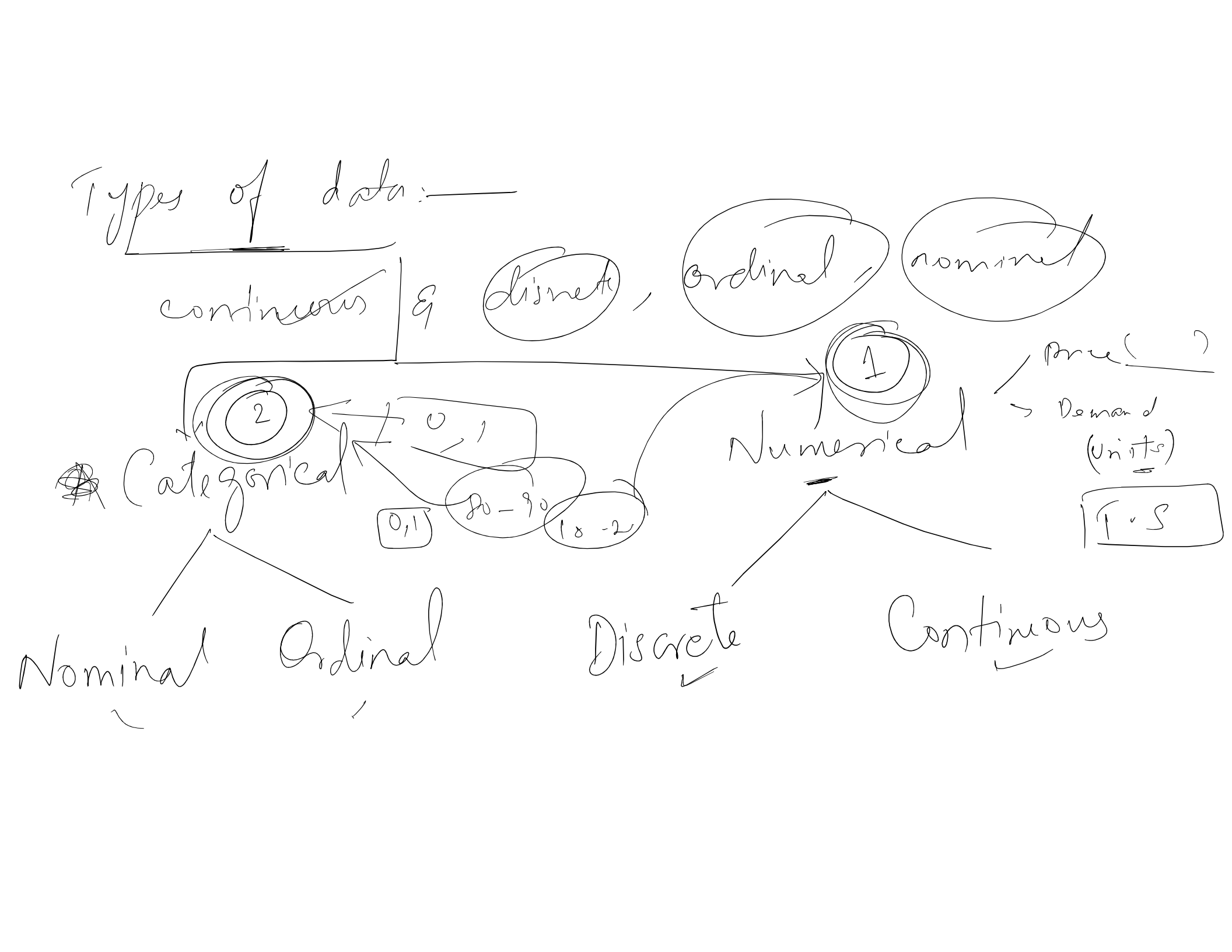
Agenda:

↳ Recap on Statistics

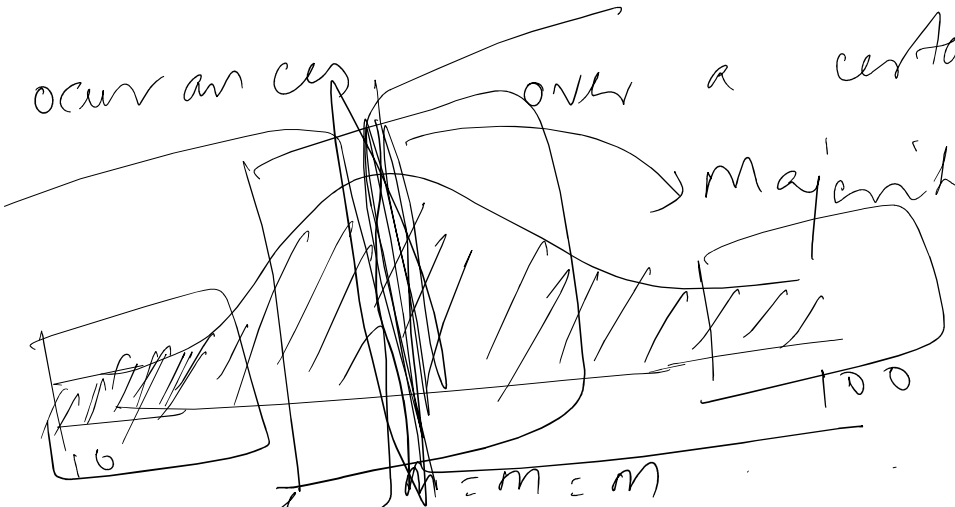
↳ Pre-processing


Types of data :—
continuous & discrete, ordinal, nominal
1
2
Categorical
0,1
20 - 90
18 - 2
0,1
Numerical
Price ( )
→ Demand (Units)
1.5
Nominal
Ordinal
Discrete
Continuous

Frequency :-

- # of occurances over a certain period

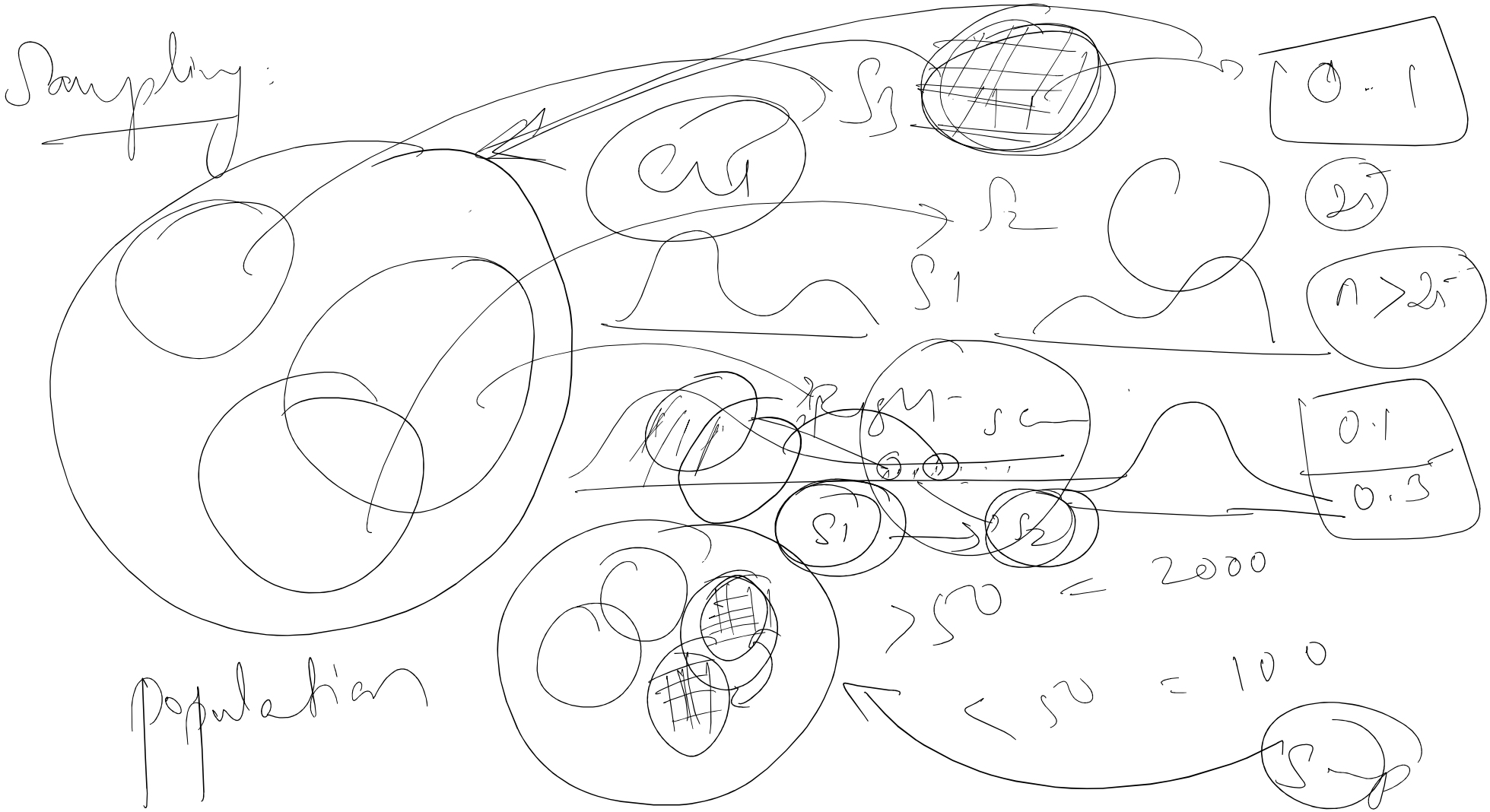
-



=> Students performance

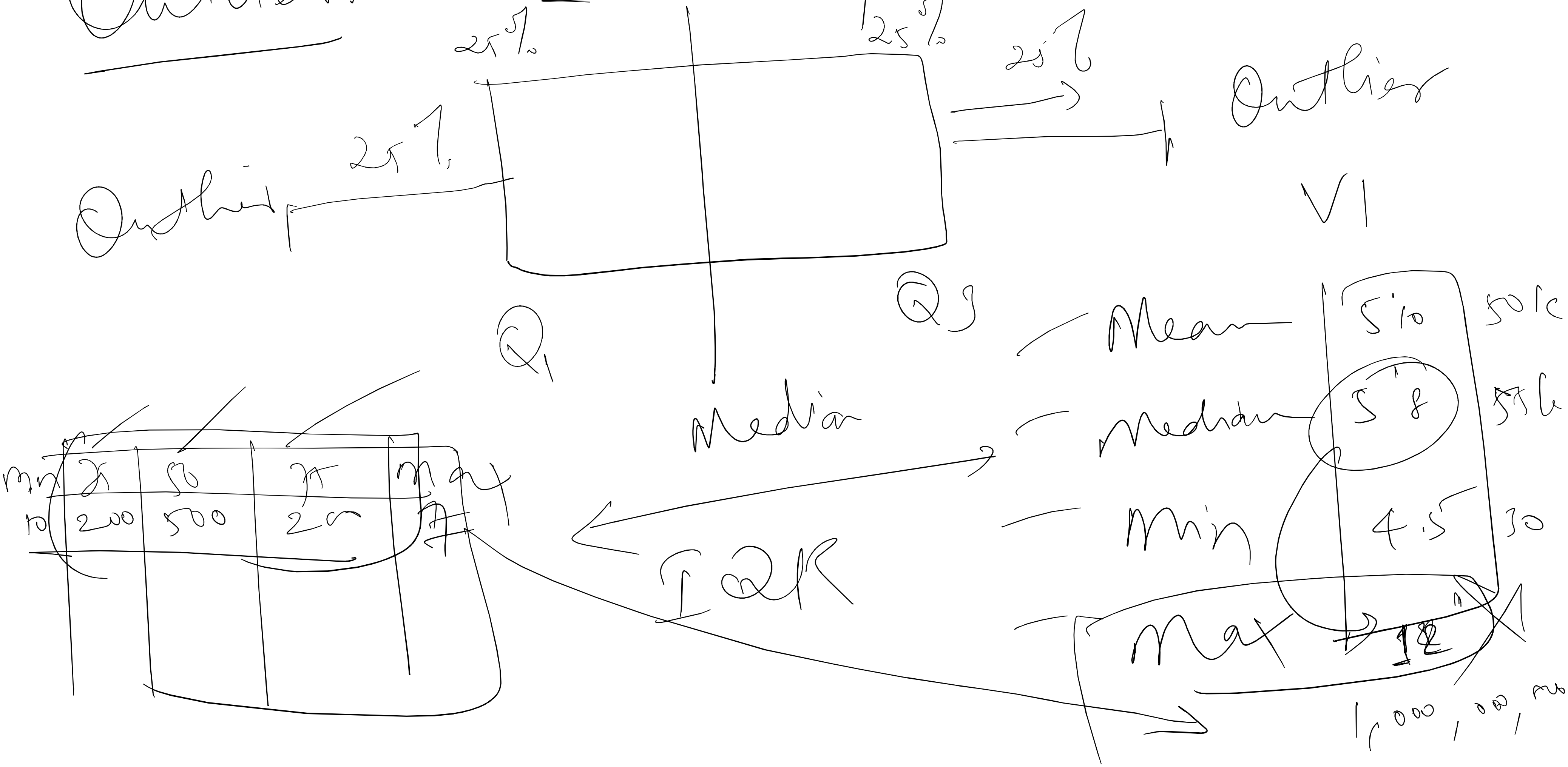
=> Normal distribution = Symmetrical


Central Limit Theorem
n>25
n>2
Calculate mean
S1
S2
S3
S4
S5
S100
Population mean
Mean


Sampling
Population
0 - 1
n > 2
0.1
0.3
< 50 = 100

Outlier:

Box - Whisker plot



Pandas – profiling: ( )

Measure of Dispersion:

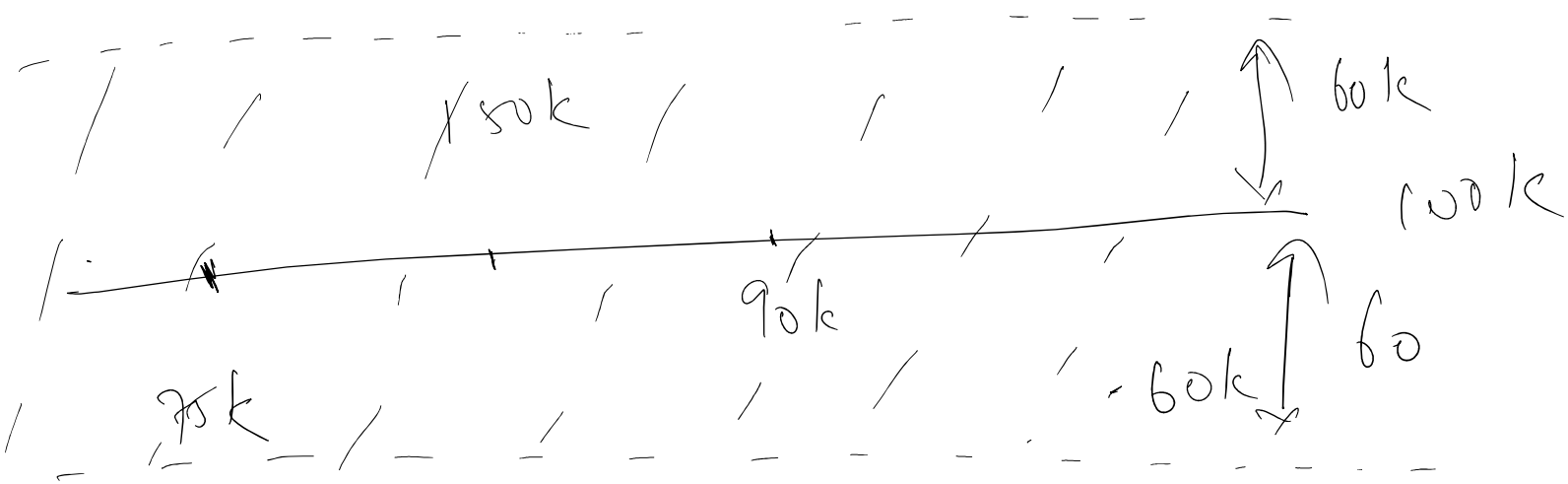
- Standard deviation
- Range (max – min)

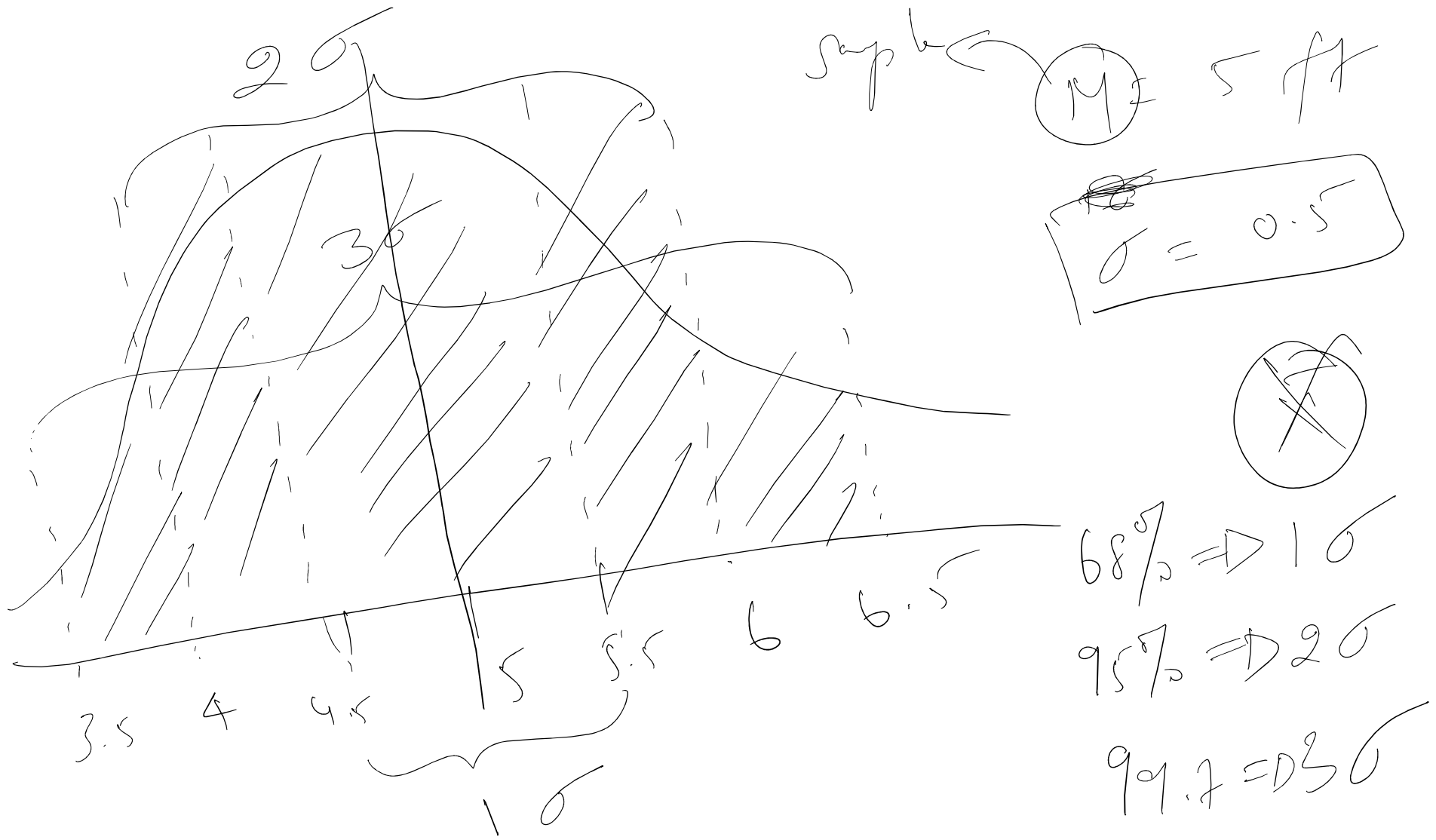
(3, 7, 9, 10, 11, 15)

15 – 3 = 12

Standard Deviation: ~20k

68%



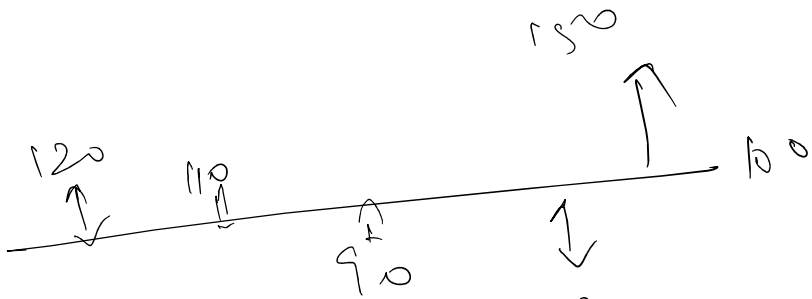

2σ
3σ
sample
M = 5
σ = 0.5
3.5
4
4.5
5
5.5
6
6.5
1σ
68% => 1σ
95% => 2σ
99.7 => 3σ

Coeff of Variation: (CV)

$$CV = \frac{\text{Std}}{\text{Mean}} \times 100$$

$$0 < CV < 1$$

$$\sqrt{V} = S.D$$



$$\frac{(20)^2 + (10)^2 + (-10)^2 + (-40)^2 + (50)^2}{5}$$

$$= \text{Variance} \Rightarrow \sqrt{1200} = S.D$$

Pen :— (24) CY2022 (24) CY2023 CY2024

J F M ... D J F M ... D J F M

Mumbai (5m) ← 1k 2k X X X 5k ...

Chennai (1m) = 12 + 12 = 24

Delhi (8m) =

Kolkata (3m) =

Bangalore (6m) = ✓ ✓ ✓ ✓ ✓