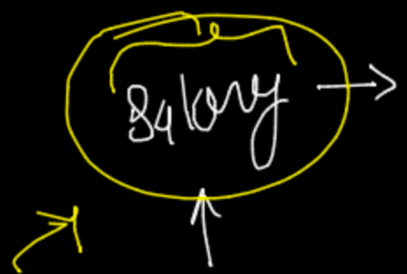
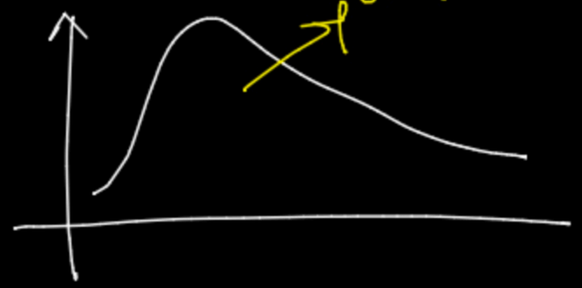
Sampling distribution (mean)





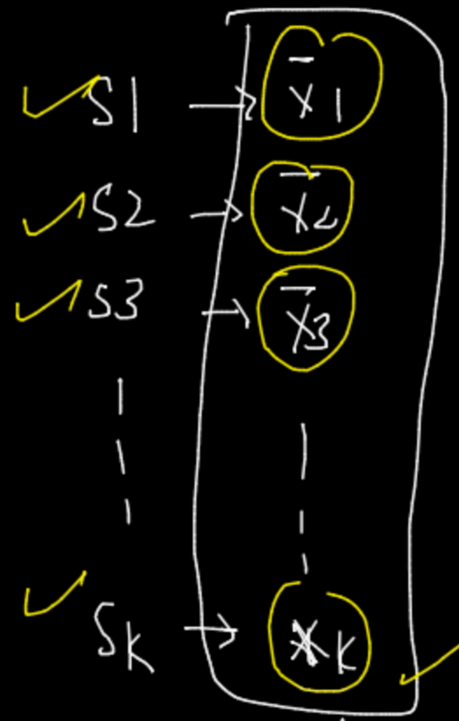
Random sample

$S1 = \{ - - - - - - - - \}$ (Let $n = 30$)

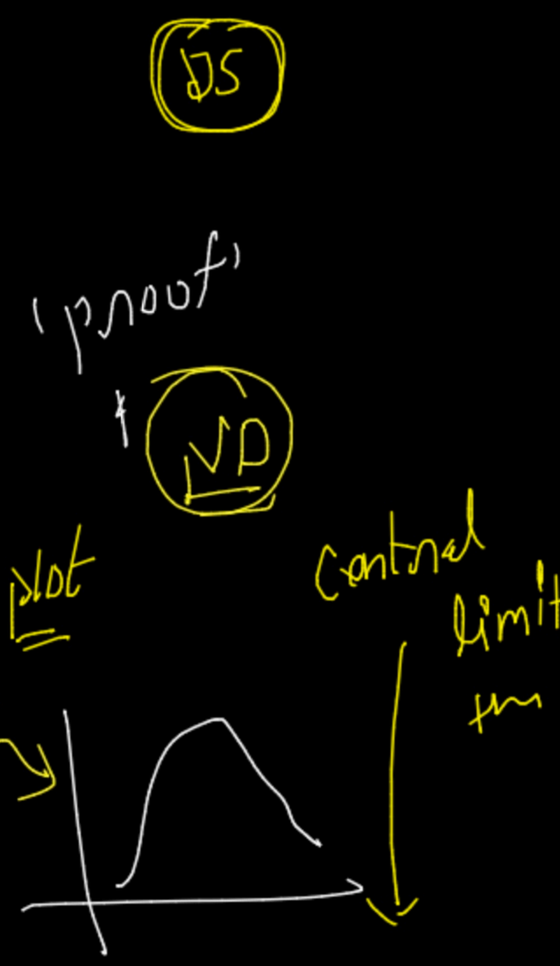
$\bar{x}$ (mean of sample)

$S2 = \{$ $\}$ (let $n = 30$)

$\bar{x}_2$ (mean of sample 2)



plot



$\bar{x}_i \rightarrow N\left(\mu, \frac{\sigma^2}{\sqrt{n}}\right)$

as $n \rightarrow \infty$

$X \sim x_1\ x_2\ x_3 - - - - x_n$

Question?

is (X) is Gaussian distribution ?

Quantile quantile plot → Q-Q plot (code)

→ statistical testing { (1) KS, (2) AD } (N2)

ths ↑ then ↑ (scipy) ← numpy

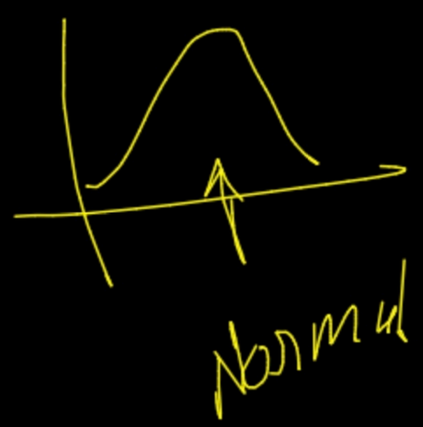
~~X~~ ~ $x_1\ x_2\ x_3\ --\ x_{500}$

(Step1) sort x_1 and compute percentile

$[x_1'\ x_2'\ --\ --\ x_{500}']$

↓ percentile

$X \Rightarrow [x_5'\ \ x_{10}'\ ------\ x_{100}^{th}]$ → plot



Step2

$Y = np.random.\widetilde{normal}(0,1)$

$Y = [\ ------\ ]$

sort

$Y \sim N(0,1)$

sort

$Y'_1 \quad Y'_2 \quad Y'_3 \quad - - - - \quad Y'_{1000}$

⇓

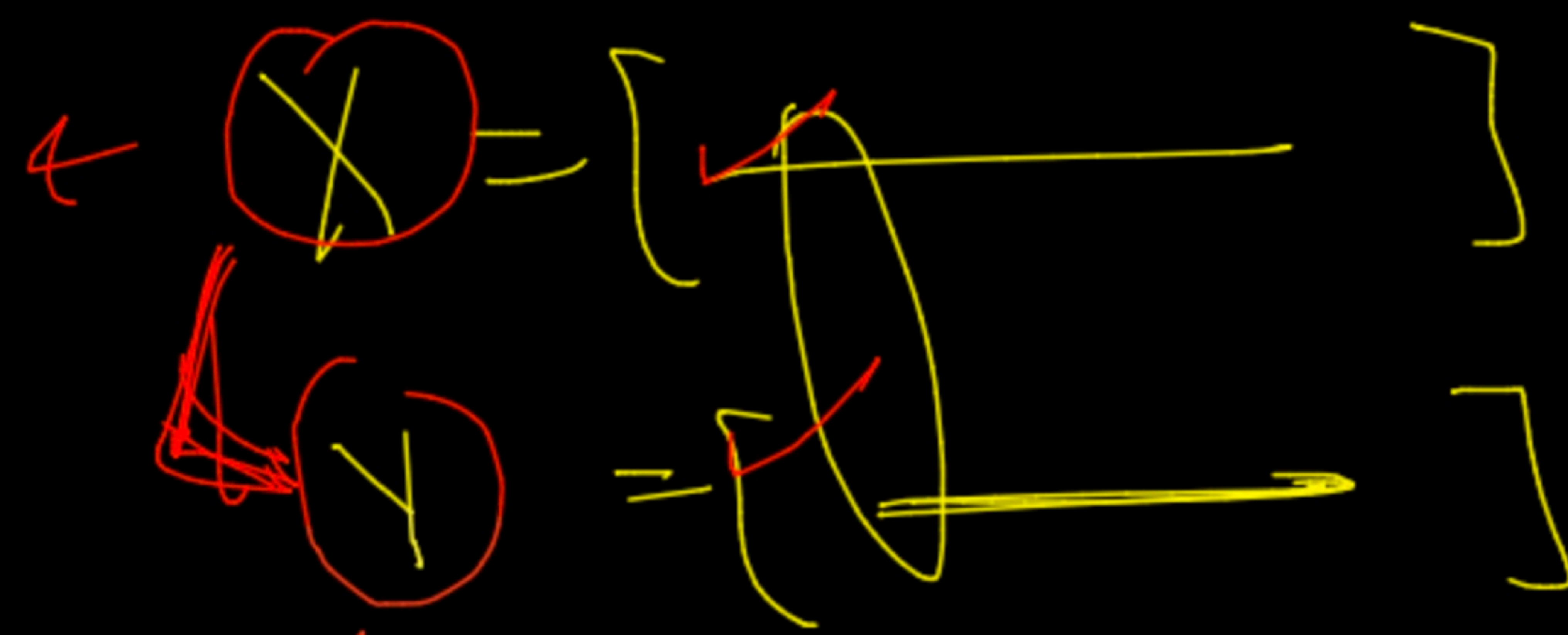
percentile

↓

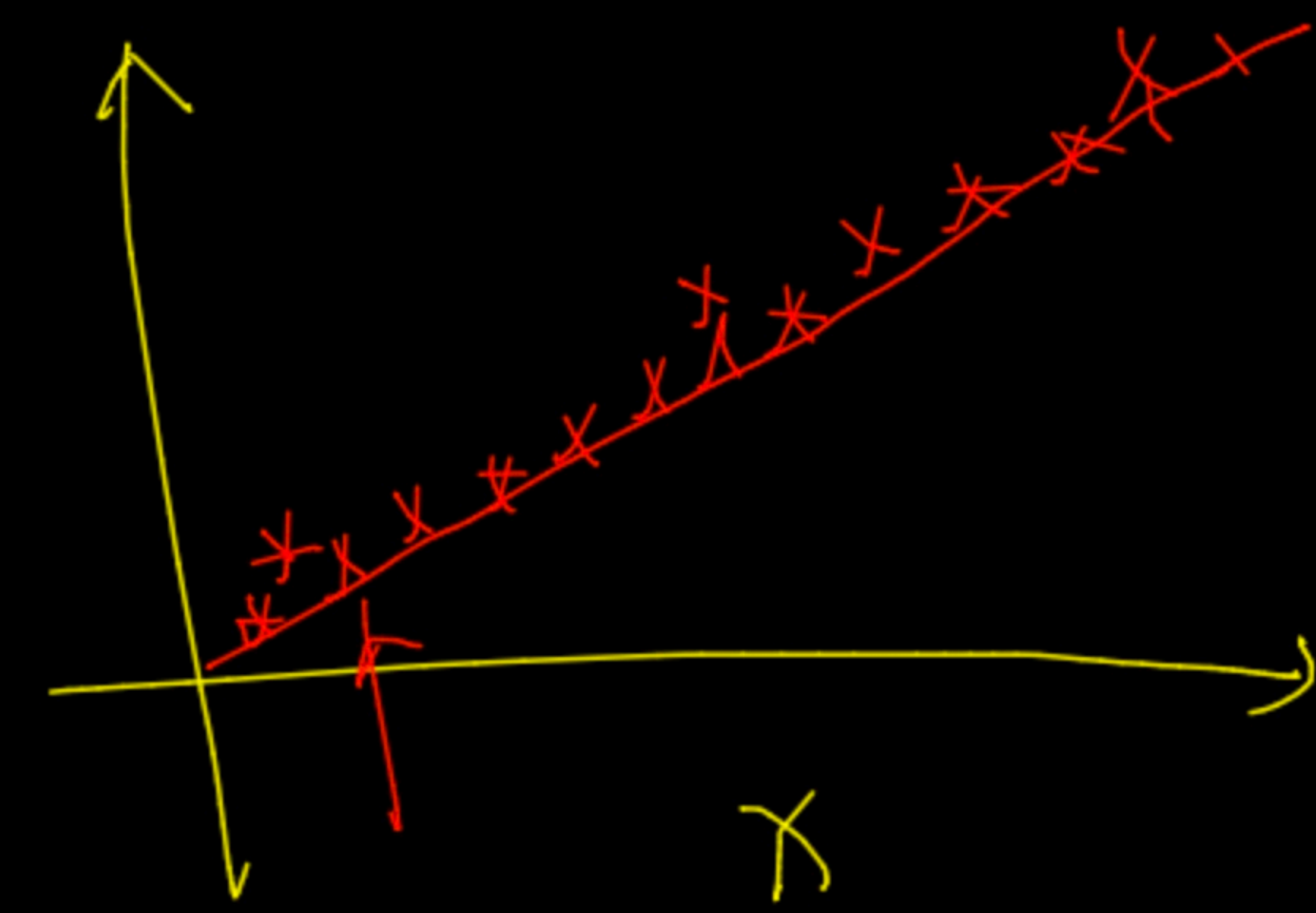
$Y_1^{(S)} \quad Y^{(L)} \longrightarrow Y_{100}$

$X_1^{(S)}$

follows

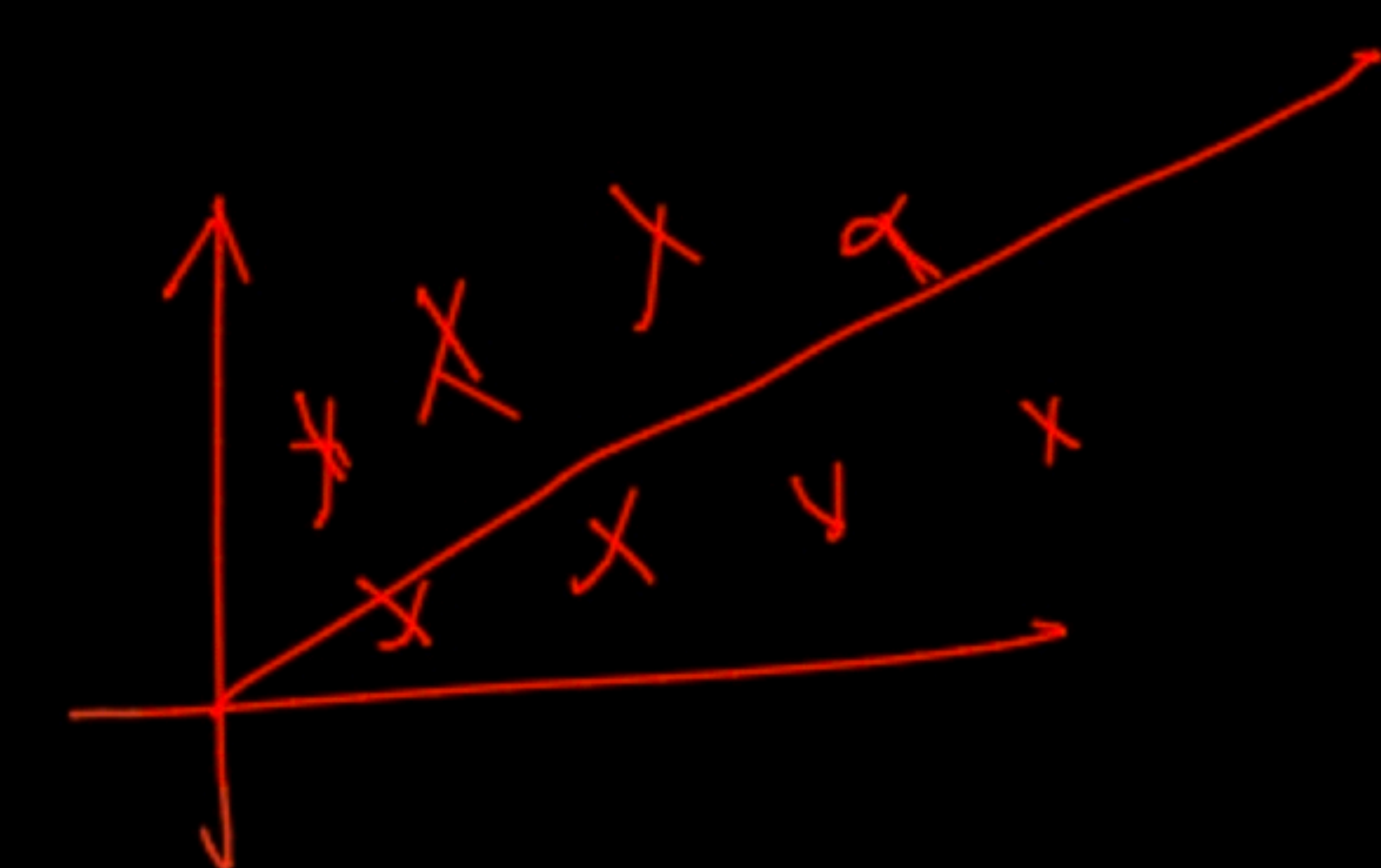


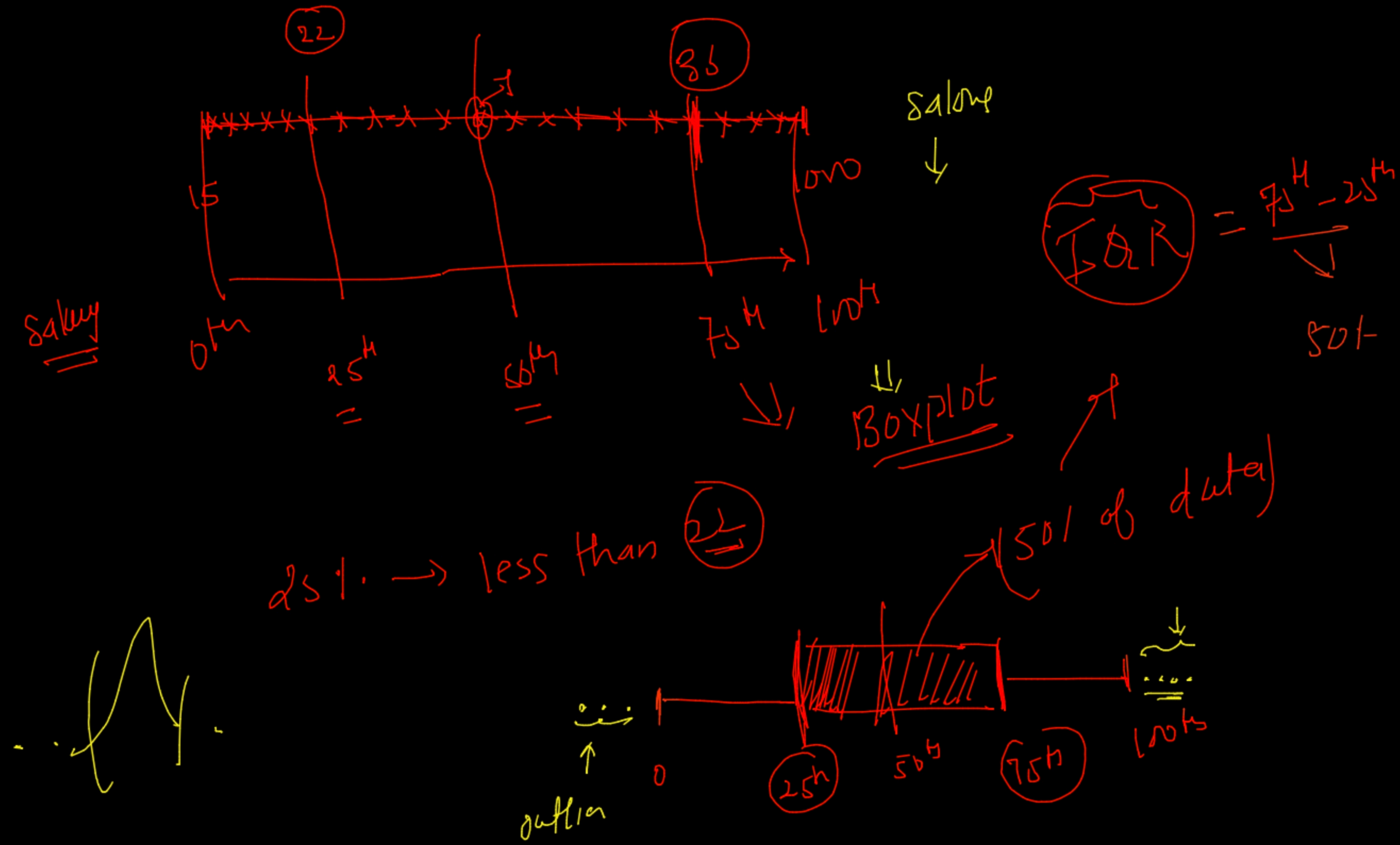
WP (case I)



$X \sim N$

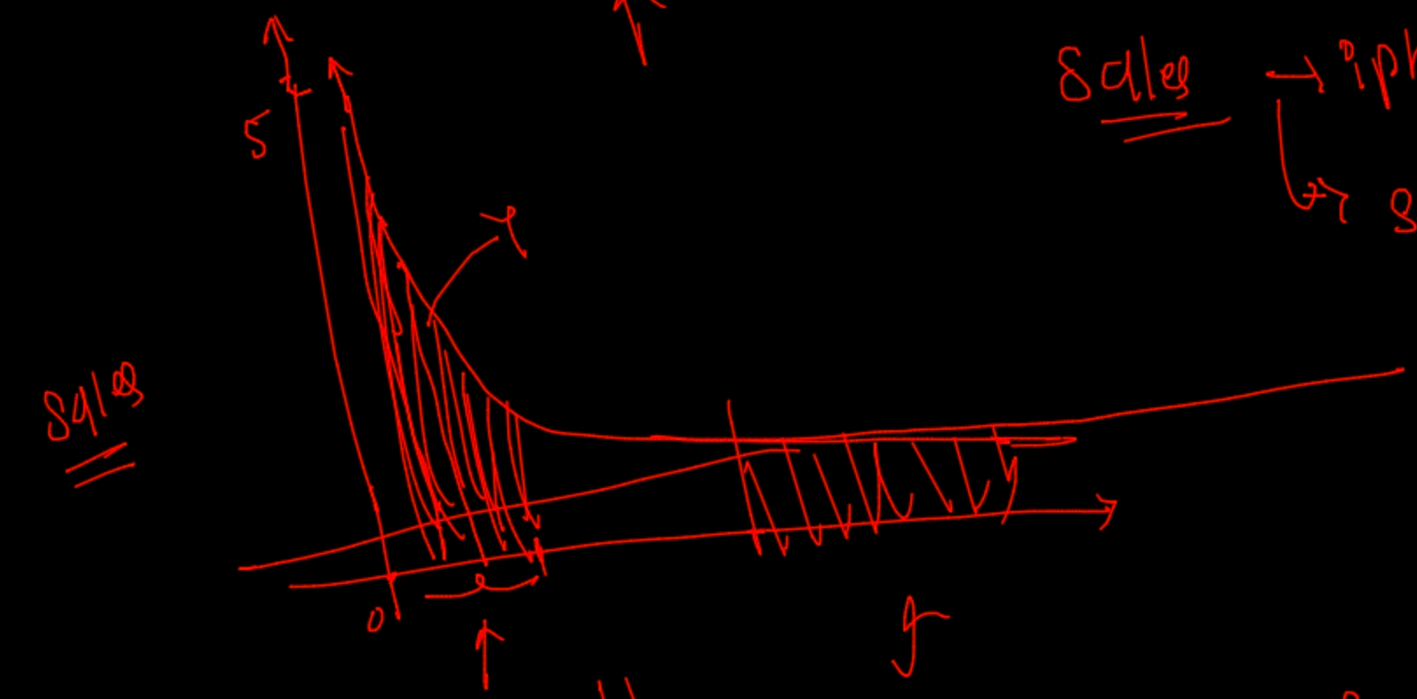
(case II)



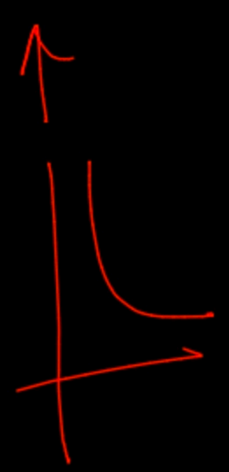

22
35
salary
15
1000
0th
25th
50th
75th
100th
Boxplot
IQR = 75th − 25th
50%
25% → less than 22
50% of data
0
25th
50th
75th
100th
outlier

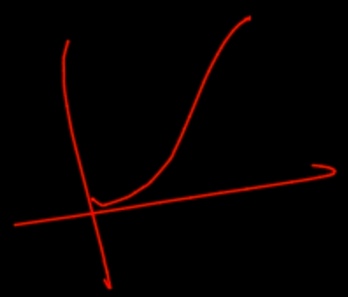
A new page has been created.

Power law (Pareto distribution) (80 - 20 Rule)



Sales → iphon (100K) → (2)
↳ samsung (10K) →

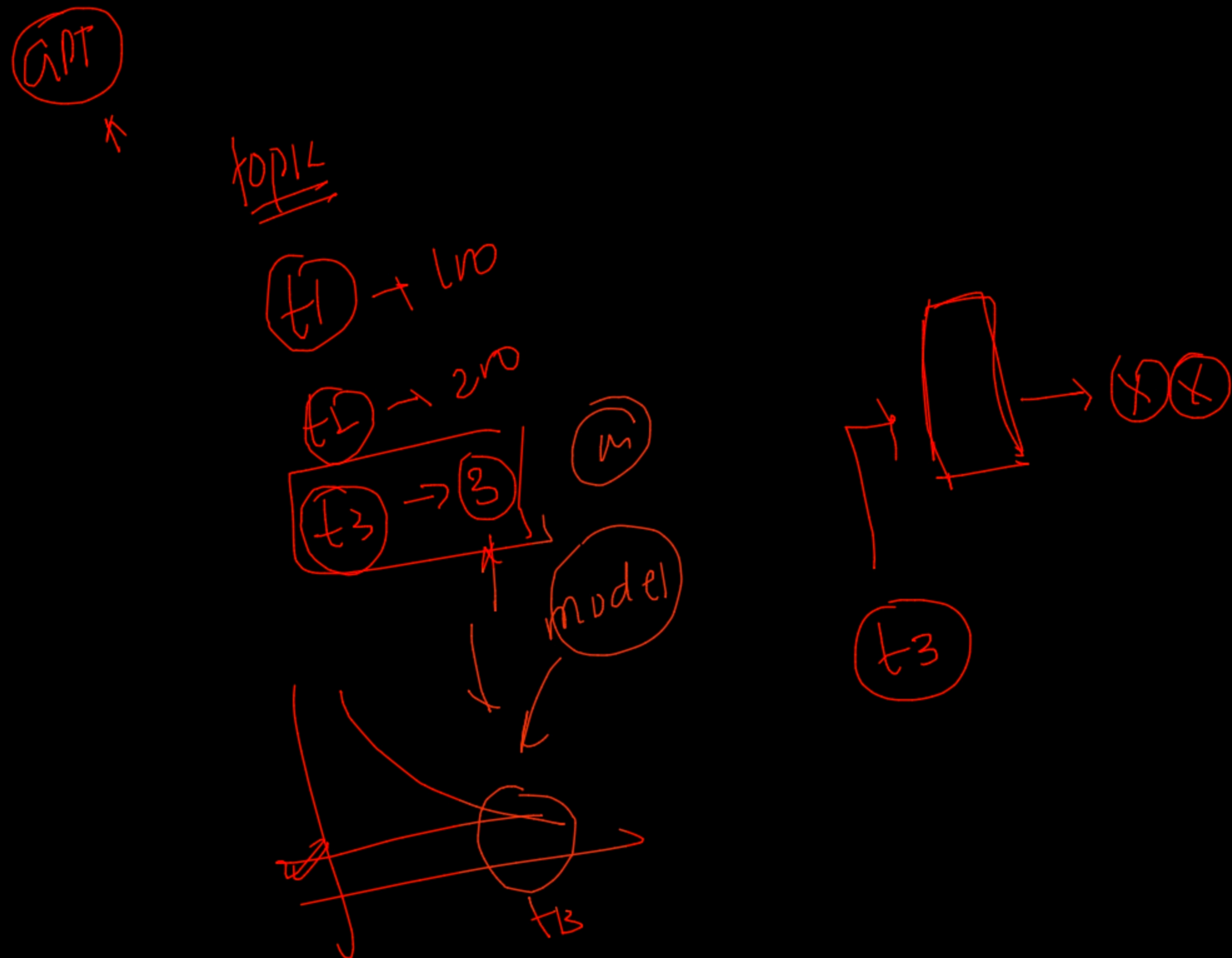




(M2)

80% time you will find things

20% popular


GPT
topic
t1
t2
t3
3
m
model
t3
t3

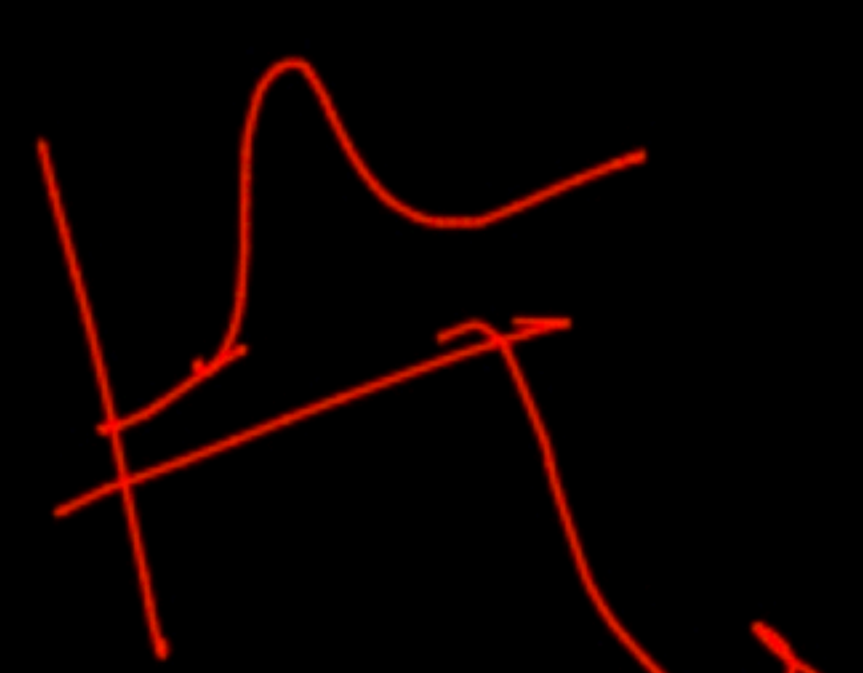
Power - transformation

⇓

(Box-Cox transformation) → SciPy

Pareto ↓

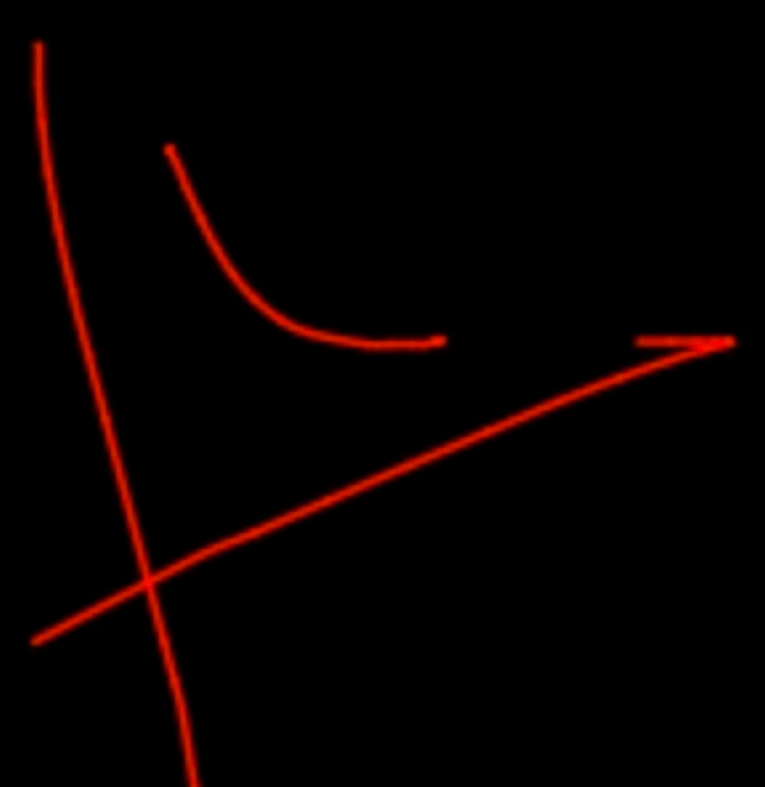
$(x_1, x_2, \ldots, x_n)$



(X) → Gaussian

↑

log



21 ✓
23 ✓

(1) X → X ?

Box-Cox(λ)

↓

(λ = 26)

Box-Cox(X) = lamda(λ)

$$y_i = \frac{x_i^{\lambda} - 1}{\lambda}, \text{ if } \lambda \neq 0$$

9 / 10

$\log(x_i)$

if $x - 0$