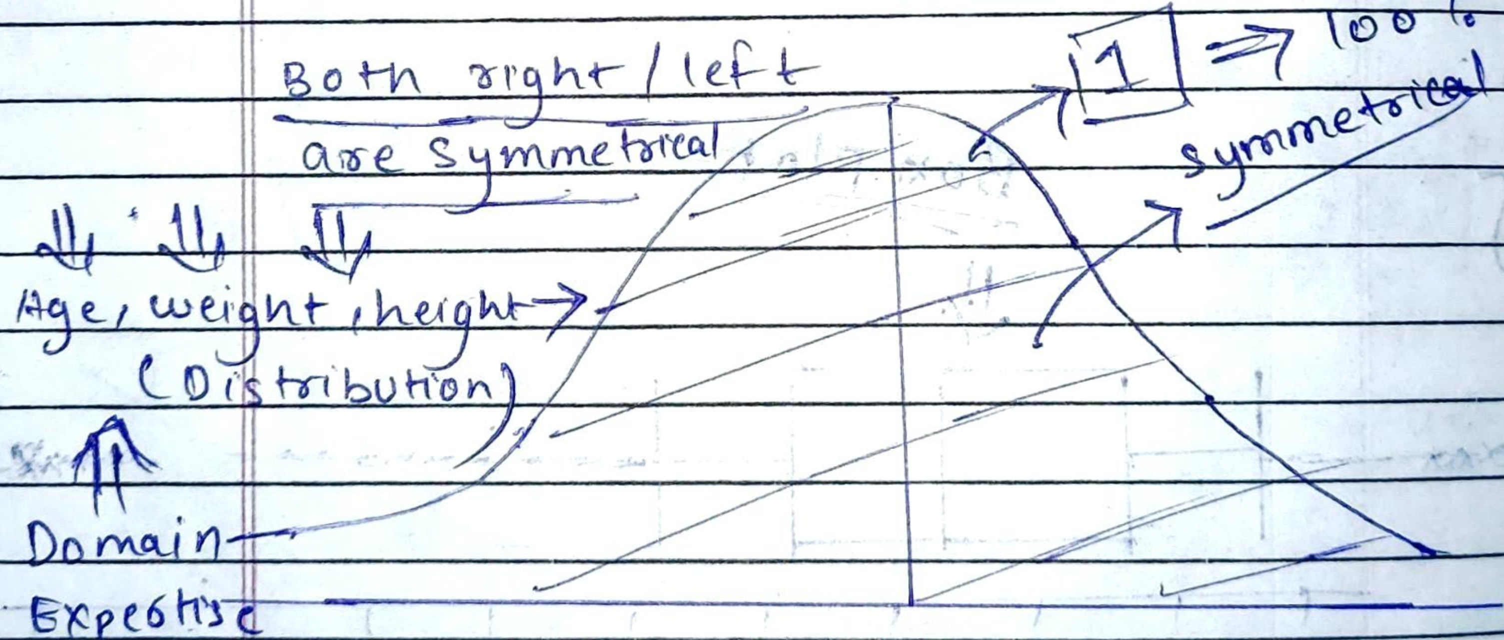
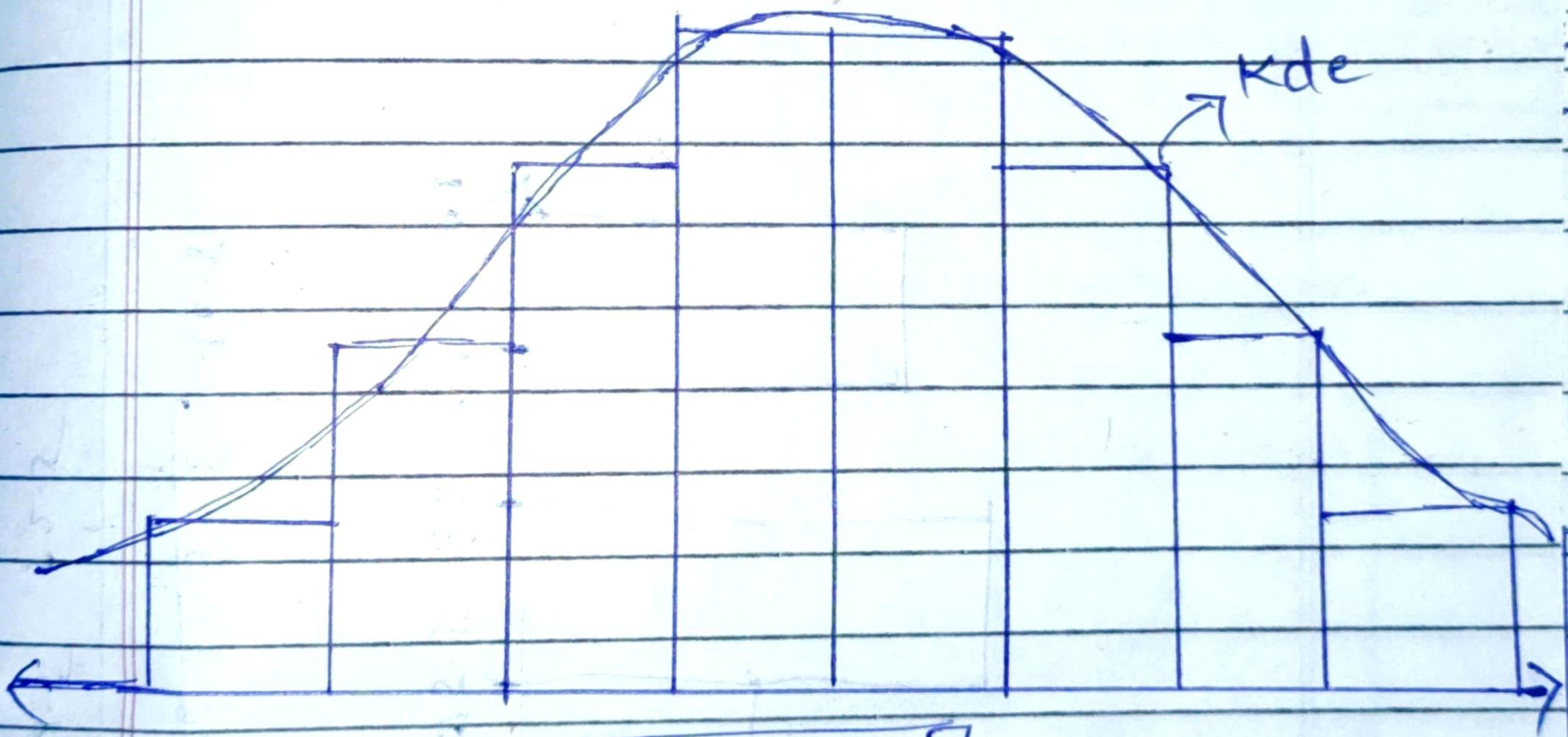
Day 3 (Statistics)

1. Normal Distribution
2. standard Normal Distribution
3. Z - score

* Gaussian / Normal Distribution



(Kde) Kernal density estimation



Histogram

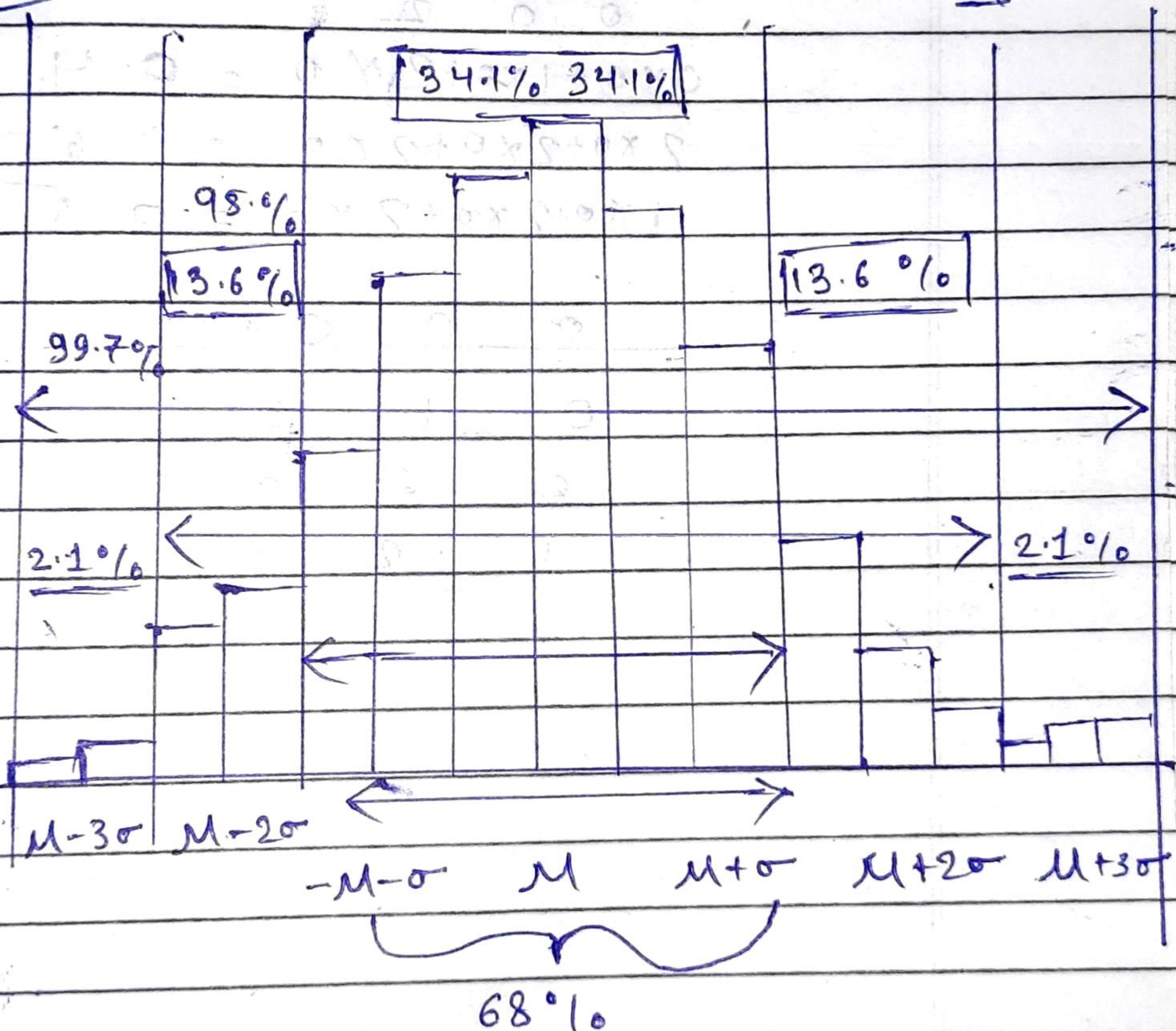
IRIS Dataset

Petal length, Sepal length, Petal width, Sepal width

⇓

Gaussian Distribution.

* Emperial Rule of Normal Distribution



Gaussian / Normal Distribution

⇓

Assumption of Data.

Page No.

Date :

(SND) Standard Normal Distribution :-

$X \approx$ Gaussian Distribution (μ, σ)

$\Downarrow \qquad \Downarrow \qquad$ Z-score $= \frac{x_i - \mu}{\sigma}$

$Y \approx$ SND $(\mu = 0, \sigma = 1)$

$X = \{1, 2, 3, 4, 5\}$

$\downarrow$

$\mu = 3$

$\sigma = 1.41$

$$\boxed{\text{Z-score} = \frac{x_i - \mu}{\sigma}} = \frac{x_i - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$\boxed{\frac{\sigma}{\sqrt{n}}}$ $\Rightarrow$ Standard Error $\Rightarrow$ Inferential stats.

$$\text{Z-score} = \boxed{\frac{x_i - \mu}{\sigma}} \Leftarrow$$

$X = \{1, 2, 3, 4, 5\}$

Page No.

Date :

$$= \frac{1-3}{1.414} = -1.414$$

$$= \frac{2-3}{1.414} =$$

$$\frac{4-3}{1.414} = \frac{1}{1.414}$$

$$z\text{-score} = \boxed{\frac{x_i - \mu}{\sigma}} \Leftarrow$$

$$x = \{1, 2, 3, 4, 5\}$$

$\mu = 3$ $\sigma = 1.414$

$$y = \{-1.414, -0.707, 0, 0.707, 1.1414\}$$

process $\boxed{\text{standardization}} \Rightarrow \mu = 0$ & $\sigma = 1$

~~[standard Deviation] ⇒ μ = 0 & σ = 1~~

$\mu = 0$

$\sigma = 1$

⇒

(year) Age	(kg) weight	(cm) Height
24	72	150
26	78	160
32	84	165
33	92	170
34	87	150
28	83	180
29	80	175

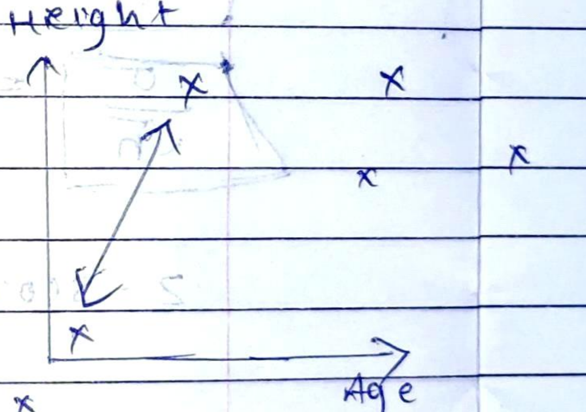
Machine learning Alg

Math equation

⇓

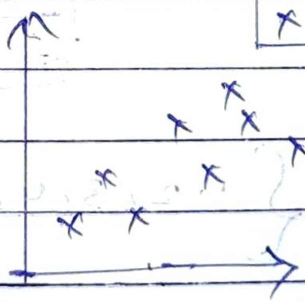
Algorithm ⇒ Mathematical ~~model~~

Mathematical calculation Time ↑ ↑↑



$\boxed{0-1}$ $\boxed{\text{Same Scale}}$

feature Scaling



$$\boxed{\frac{X_i - \mu}{\sigma}}$$

Normalization

⇓

[0 – 1] [–1 – 1]

[0 – 8]

[0 – 4]

standardization {z-score}

$\mu = 0, \sigma = 1$

[–3 ⟷ 3]

Feature Scaling :

Normalization [lower scale ⟷ Higher scale]

(1) Min Max Scaler [0 - 1]

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \xRightarrow{\text{Apply}} \text{Deep learning}$$

$$= \frac{1 - 1}{5 - 1} = 0 \qquad \frac{2 - 1}{5 - 1} = \frac{1}{4}$$

$$\frac{3 - 1}{5 - 1} = \frac{2}{4} \qquad \frac{4 - 1}{5 - 1} = \frac{3}{4}$$

$$\frac{5 - 1}{5 - 1} = \frac{4}{4} = 1$$

x	y
1	0
2	0.25
3	0.5
4	0.75
5	1

$\boxed{0 \text{ to } 1}$

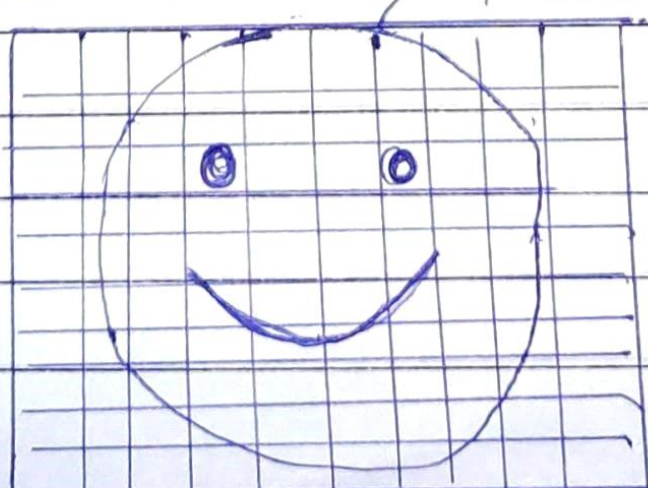
⇓

Apply

Deep learning :

$\boxed{0-255} \rightarrow \boxed{0-1}$

Pixels →



⇒ Normalization

Page No.
Date :

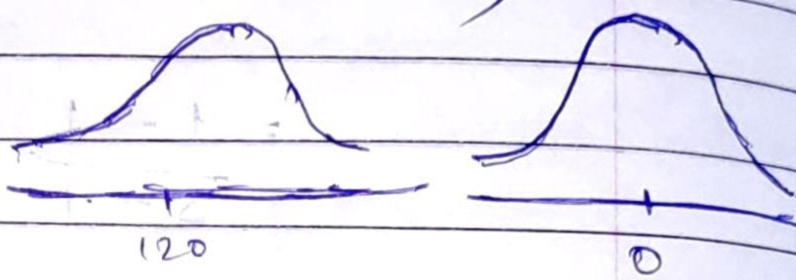
(1) standardization :

$$\boxed{z\text{-score} = \frac{x_i - \mu}{\sigma}}$$

$x \rightarrow$ Normal Distribution (μ, σ)

$\Downarrow$ z-score

$y \rightarrow$ SND $(\mu = 0, \sigma = 1)$



why do we do this $\longrightarrow$ Bring the feature in the same scale

(2) Normalization $[0 - 1]$

(1) Min Max Scaler $\Longleftrightarrow$ standardization

$\downarrow$ ML $\qquad \Downarrow$ ML

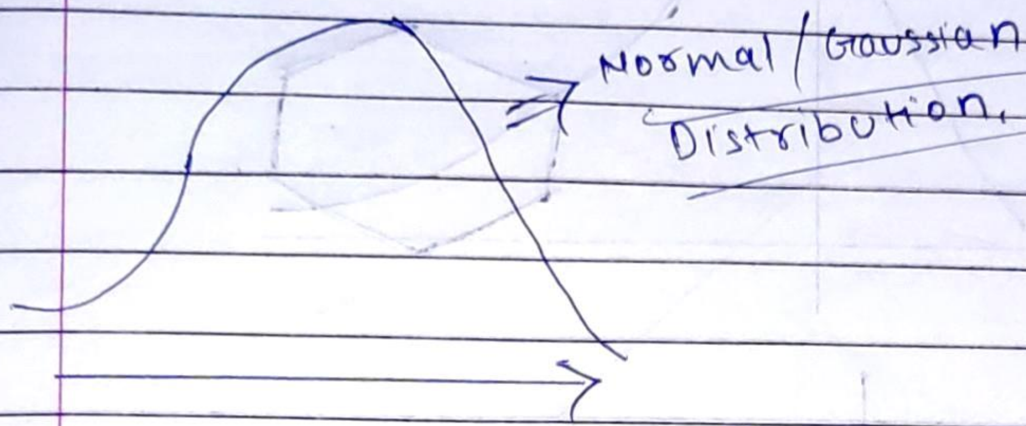
ML

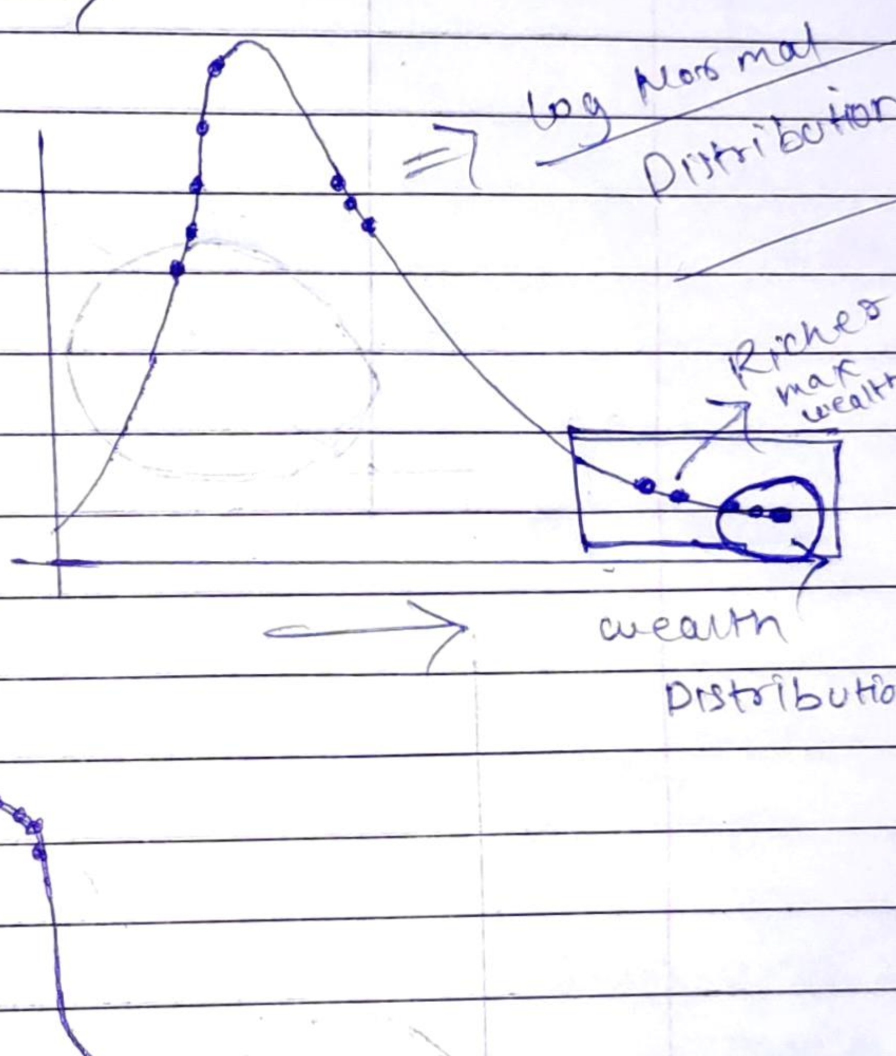
$\Downarrow$

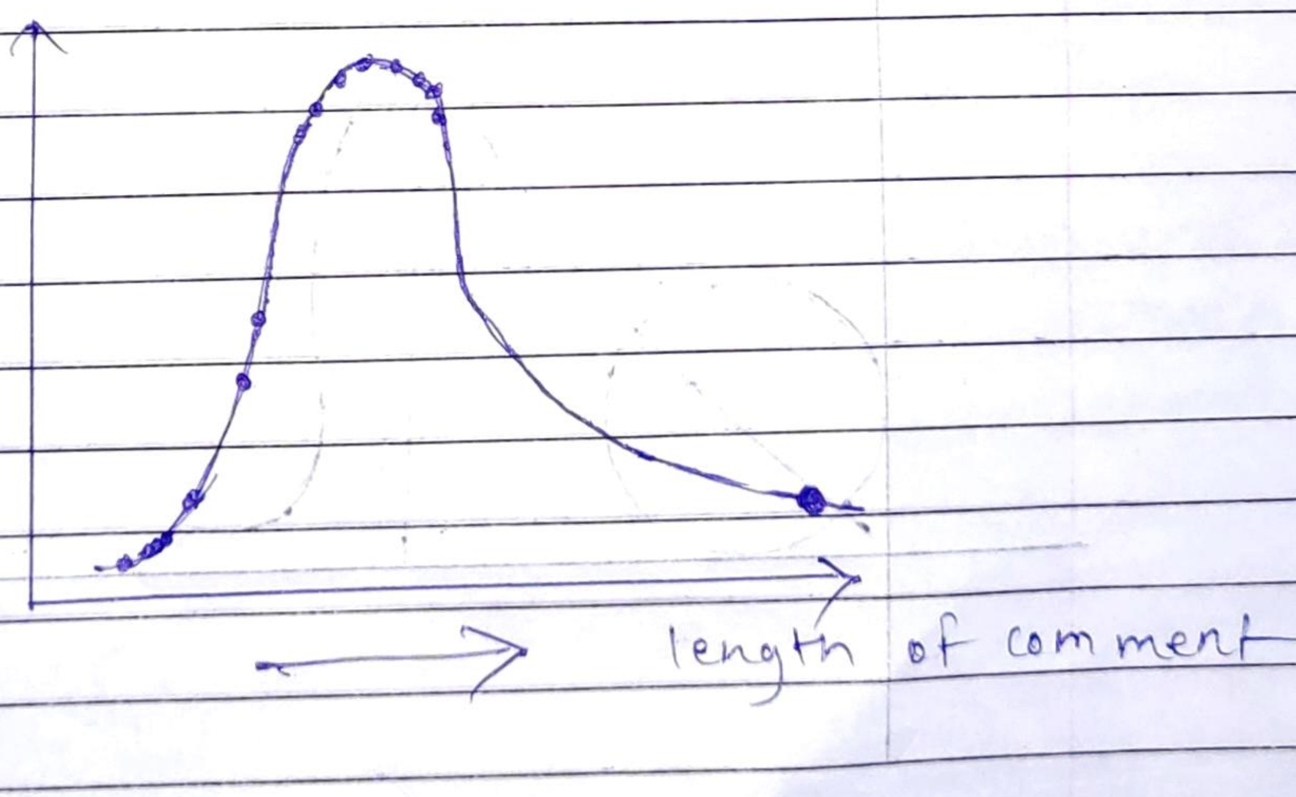
Min max scaler

Page No.
Date :

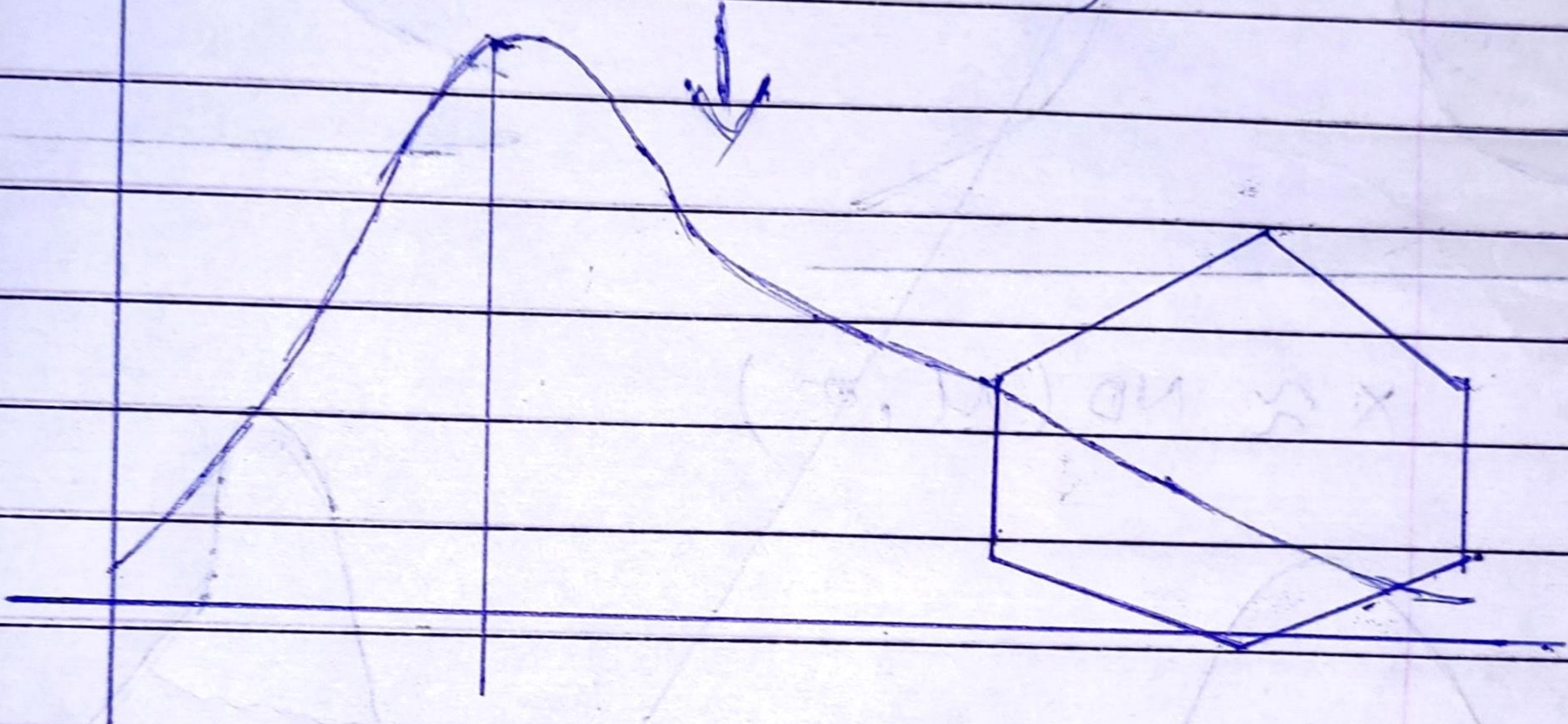
① Log Normal Distribution



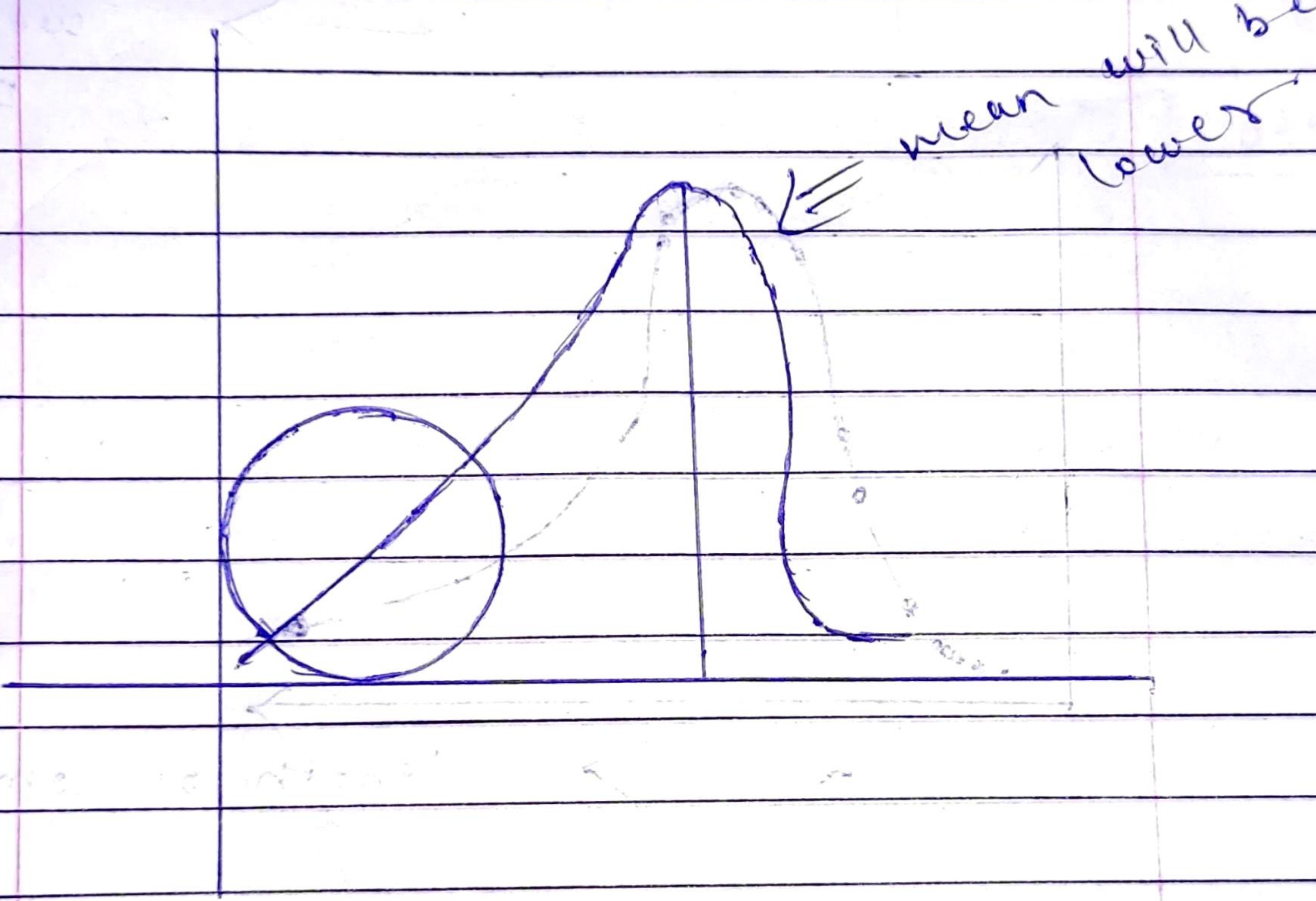




mean will be higher

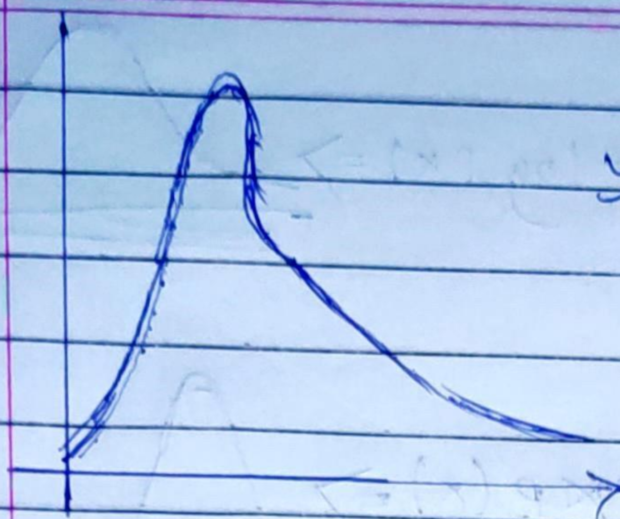


Relation of mean, median and mode.

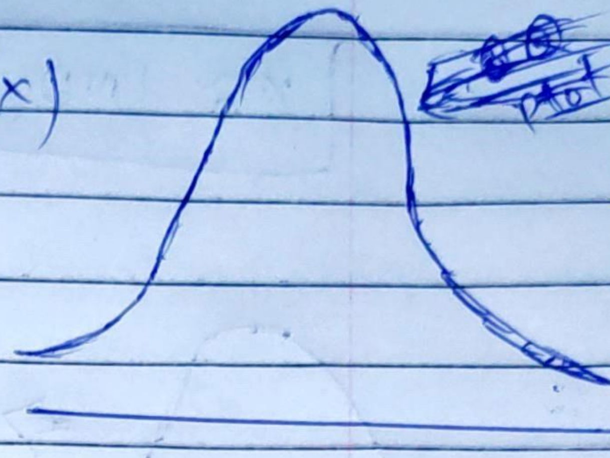


mean will be lower

Page No.
Date :



$y = \ln(x)$

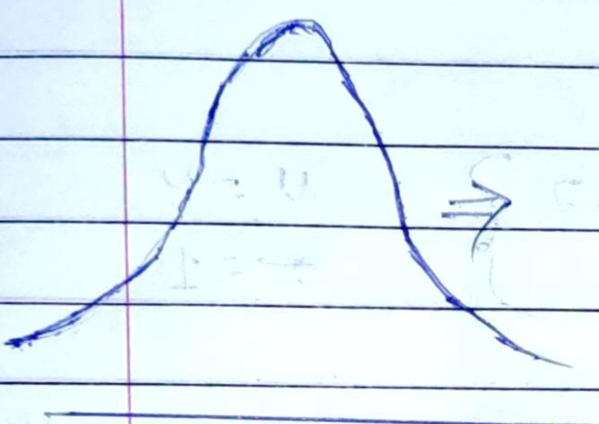


no.of Runs.

$x \simeq$ log Normal Distribution

antilog

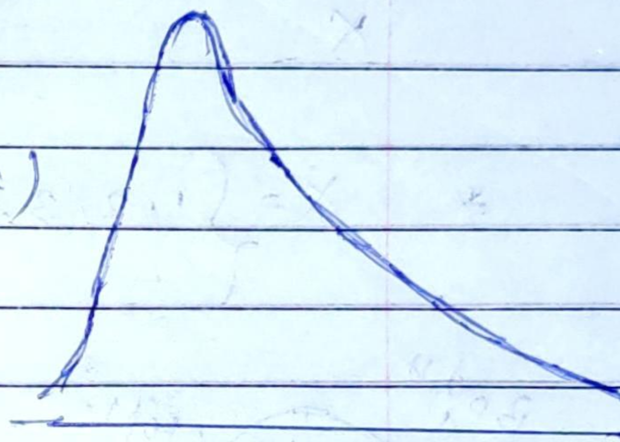
$\Downarrow$



$\Rightarrow$ $\exp(x)$

$\Downarrow$

y



$x \simeq ND(\mu, \sigma)$

normal distribution

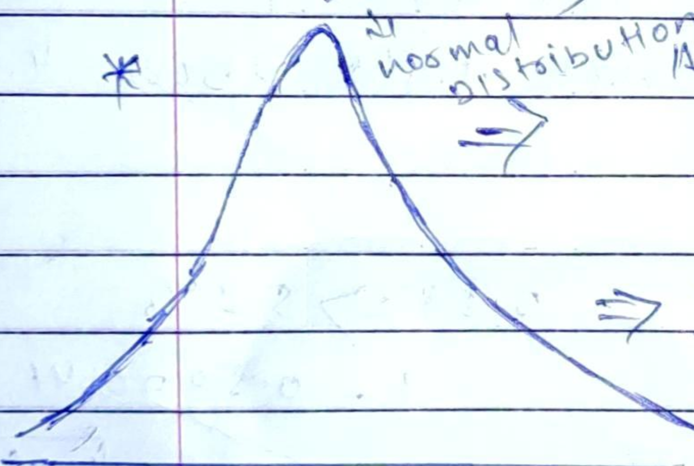
Natural log.

~~Natural log~~

$\Uparrow$

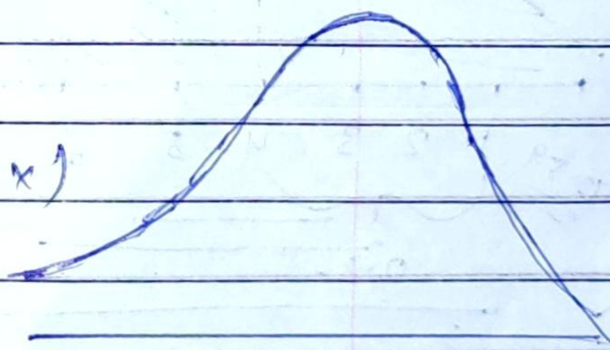
log

*



$\Rightarrow$

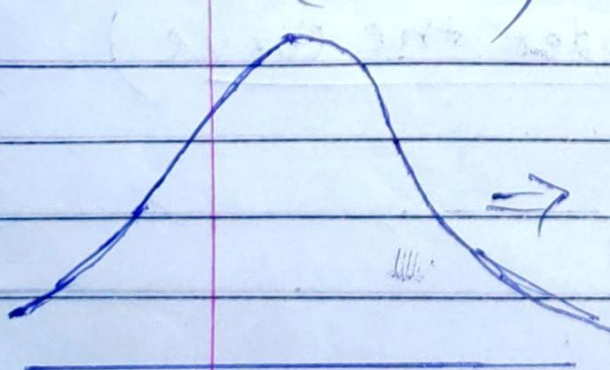
$\Rightarrow y = \ln(x)$



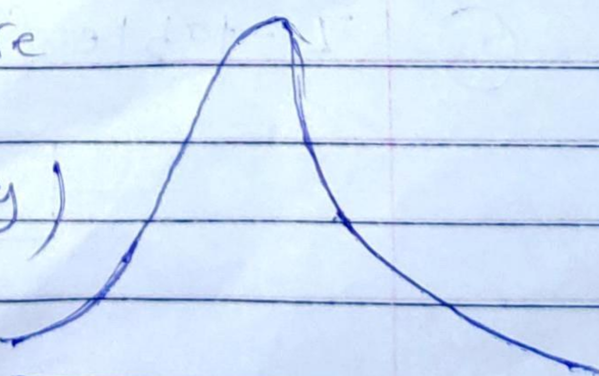
$x \simeq$ log Normal Distribution (μ, σ)

$\Downarrow$

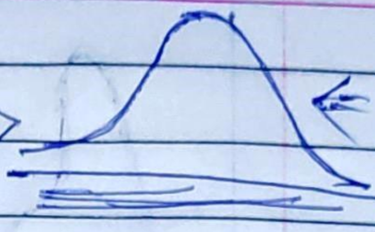
Inverse



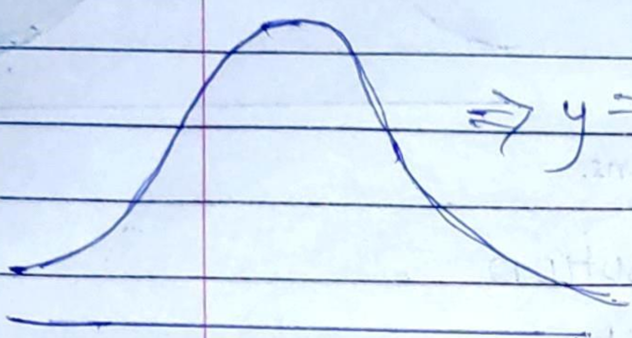
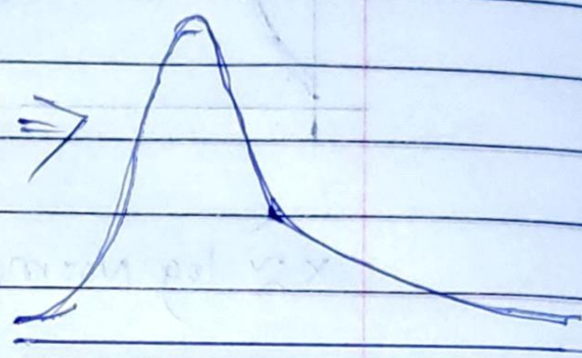
$\Rightarrow x = \exp(y)$



y

$\boxed{X \cong LND} \Rightarrow \log_e(x) \Rightarrow$  $\Leftarrow$ QQ plot

$\Downarrow$

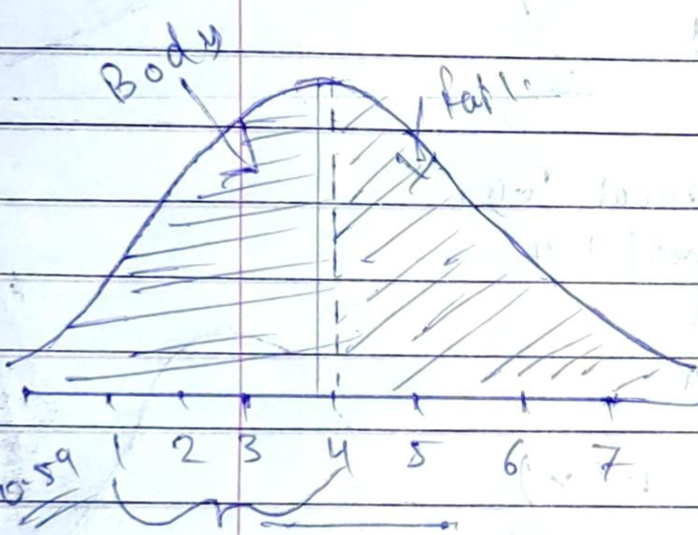
 $\Rightarrow y = \exp(x) \Rightarrow$ 

X

* $X = \{1, 2, 3, 4, 5, 6, 7\}$

$\mu = 4$

$\sigma = 1$



Question: what is the percentage of score that falls above 4.25?

$0.59 \Rightarrow 59\%$

$1 - 0.59 = 0.41$

$= \boxed{41\%}$

(1) $Z\text{-score} = \dfrac{X_i - \mu}{\sigma} = \dfrac{4.25 - 4}{1} = \boxed{0.25}$

(1) Z-table (Area under the curve)

$$Z\text{-Score} = \frac{3.75 - 4}{1} = -0.25$$