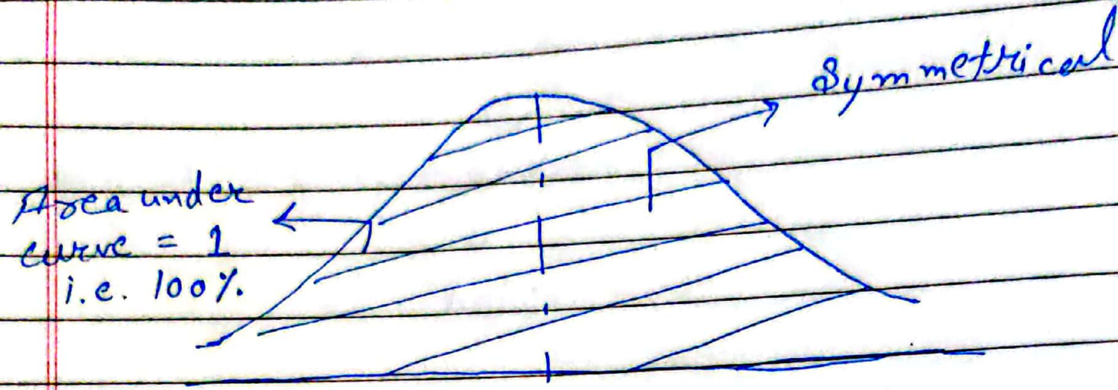
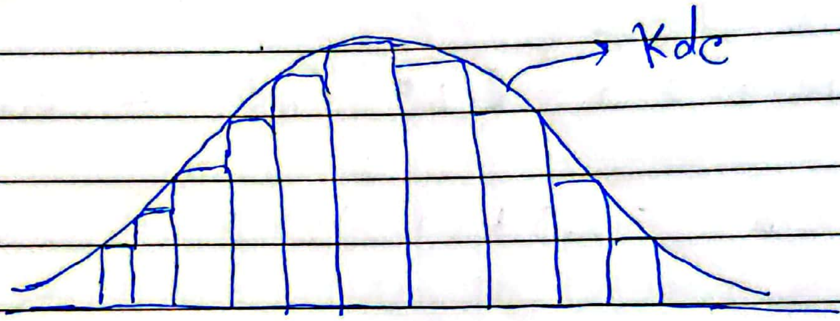
* Gaussian / Normal Distribution





Age, Weight, height

⇑ ⇒ Doctors

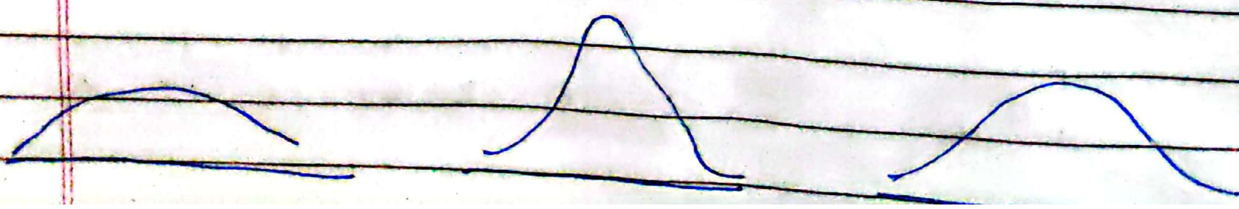
Domain Expert

$\boxed{\text{Iris Dataset}}$

⇓

~~Spe~~ Petal length, Petal width, Sepal length, Sepal width

↓ Gaussian Distribution



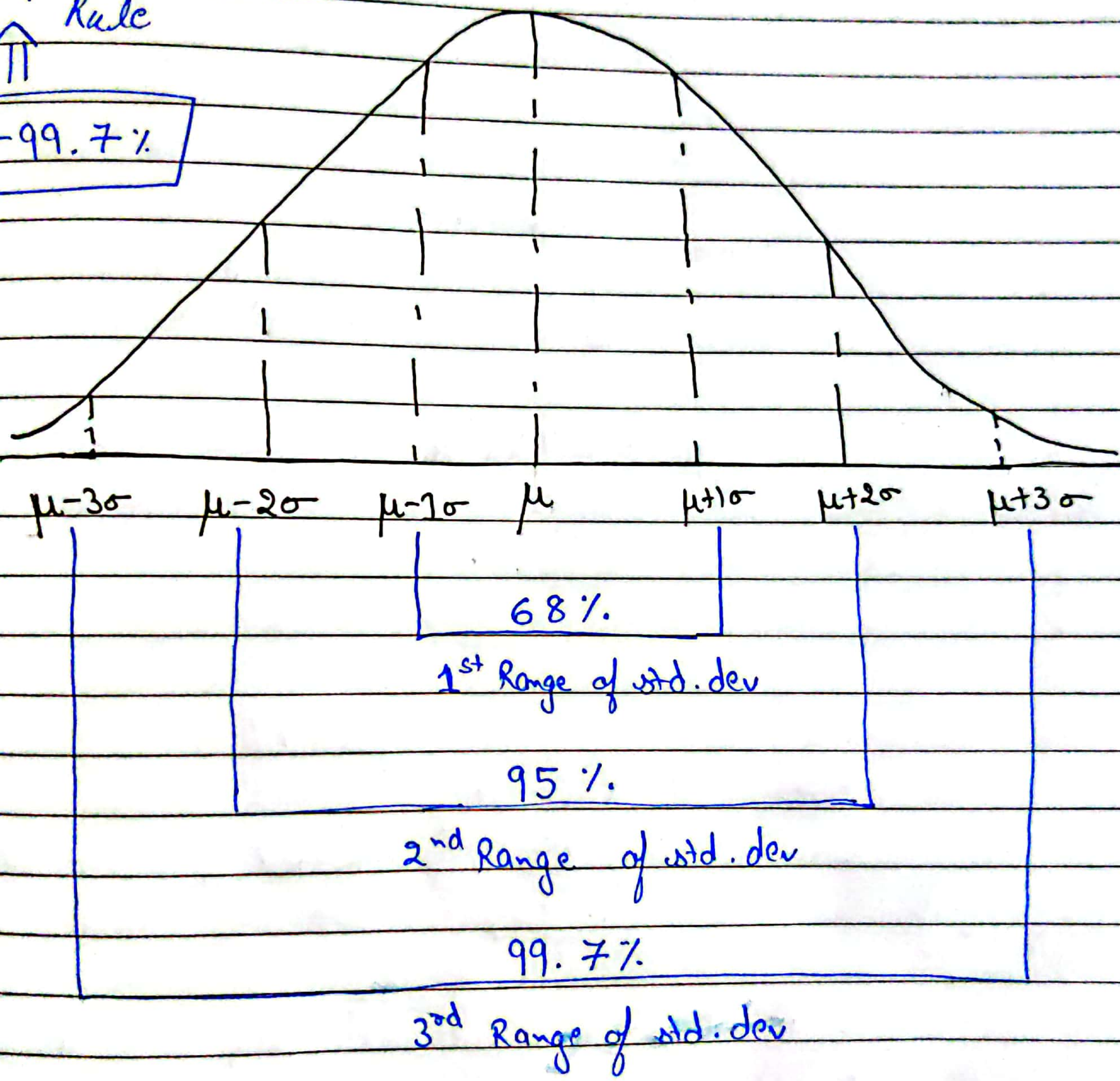
Date ______
Page ______

* Emperical Rule of Normal Distribution

Emperical ~~Formula~~ Rule

⇑

$\boxed{68-95-99.7\%}$



$\boxed{\text{Q-Q Plot}}$ ⇒ whether a distribution is Normal/Gaussian or not.

CamScanner

Date
Page

* Standard Normal Distribution

Let X is a random variable belonging to Normal Dist.

$$X \sim \text{Gaussian Distribution}\,(\mu, \sigma)$$

μ → Mean

σ → std. dev.

Z-Score

Convert X to Y

Y → std. normal distribution with $\mu = 0$ & $\sigma = 1$

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

$\boxed{n = 1}$

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

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

$\boxed{n = \text{Sample Size}}$

$n = 1$ because we will apply z-score ~~to~~ each and every value

⇓

$$\text{Z-score} = \frac{x_i - \mu}{\sigma} \quad \text{when } n = 1$$

CamScanner

$X = [1, 2, 3, 4, 5]$

$\downarrow$

$\mu = 3, \quad \sigma = 1.414$

$\downarrow$

Z-Score =

$\downarrow$

$y = [-1.414, -0.707, 0, 0.707, 1.414]$

$\downarrow$

$\mu = 0, \ \sigma = 1$

zscore (1)

$\downarrow$

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

$\frac{2-3}{1.414} = -0.707$

Q Why convert Normal Dist to SND? [Standardization]

Age (yrs)	Weight (kgs)	Height (cms)
24	72	150
26	78	160
32	84	165
33	92	170
34	87	150
28	83	180
29	80	175

$\mu = 0$
$\sigma = 1$

Here each column have different units.

Machine Learning

$\downarrow$

Algorithm $\Rightarrow$ mathematical model

$\downarrow$

maths equation

As units are different so calculation will take more time

* Normalization

Standardization { z-score }

↓

$\boxed{\mu = 0, \sigma\ 1}$

<u>Normalization</u> → we give the range like convert all values b/w 0 & 1

↓

[0 - 1]

[0 - 5] etc.

~~Normdt~~

Normalization [Lower scale to higher scale]

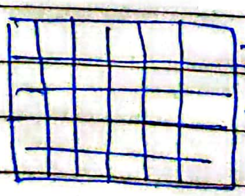
↓

① <u>Min Max Scaler</u> [0 - 1]

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

⇓

Applied in Deep Learning



→ image

→ pixels → [0-255]

↓

[0 - 1]

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

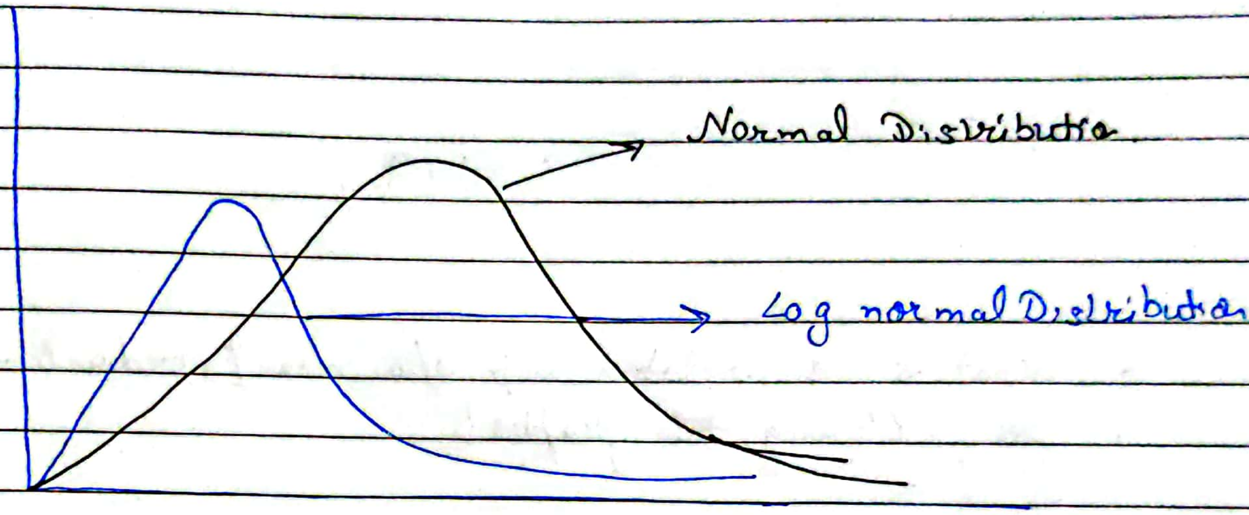
$y(1) = \frac{1-1}{5-1} = 0$

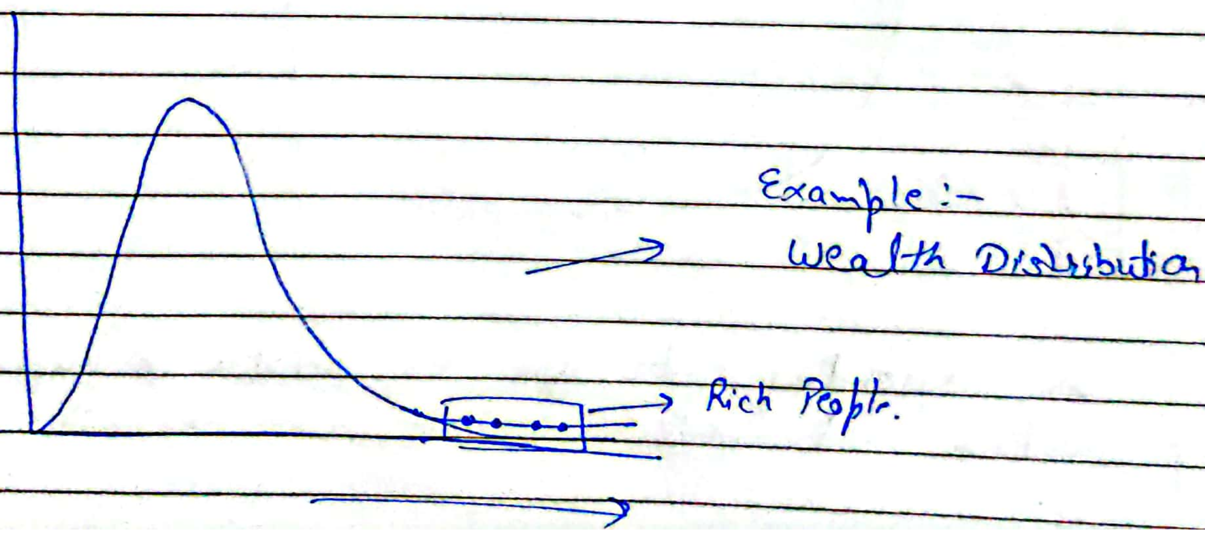
$y(2) = \frac{2-1}{5-1} = \frac{1}{4}$

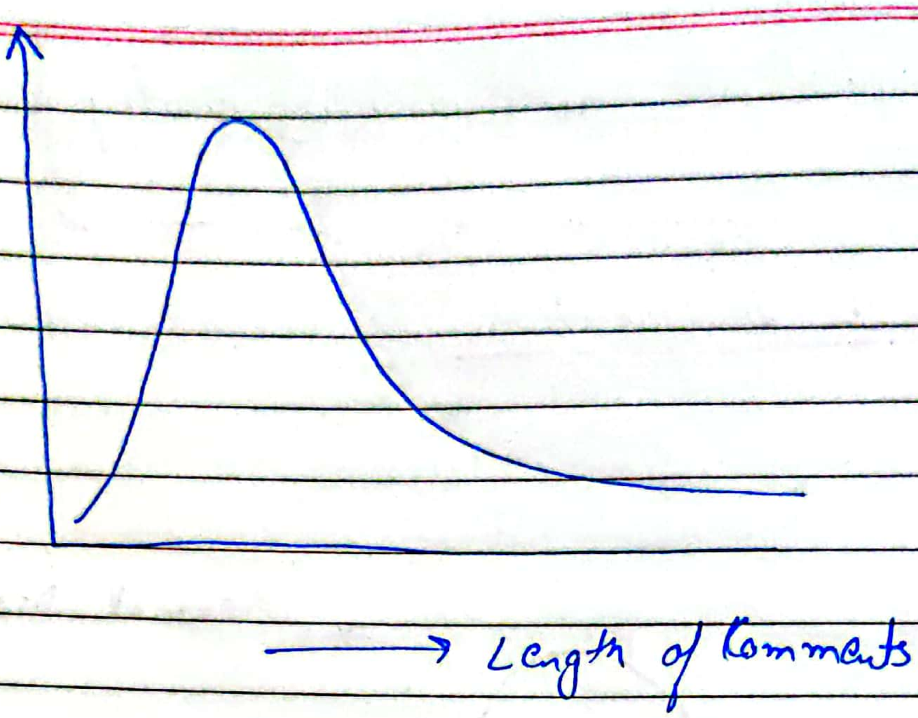
These Techniques are called Feature Scaling

---

* Log Normal Distribution

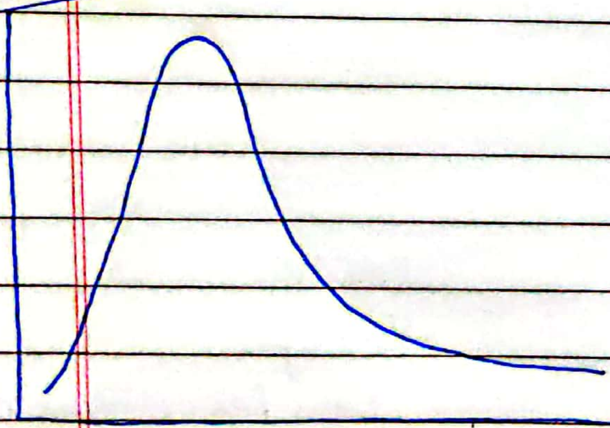


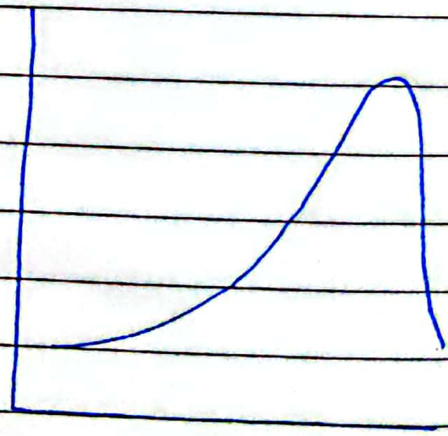




Q What is the relationship b/w mean, median & mode for following two graphs?

Interview Question





From Assending order give the relation of mean, median and mode.

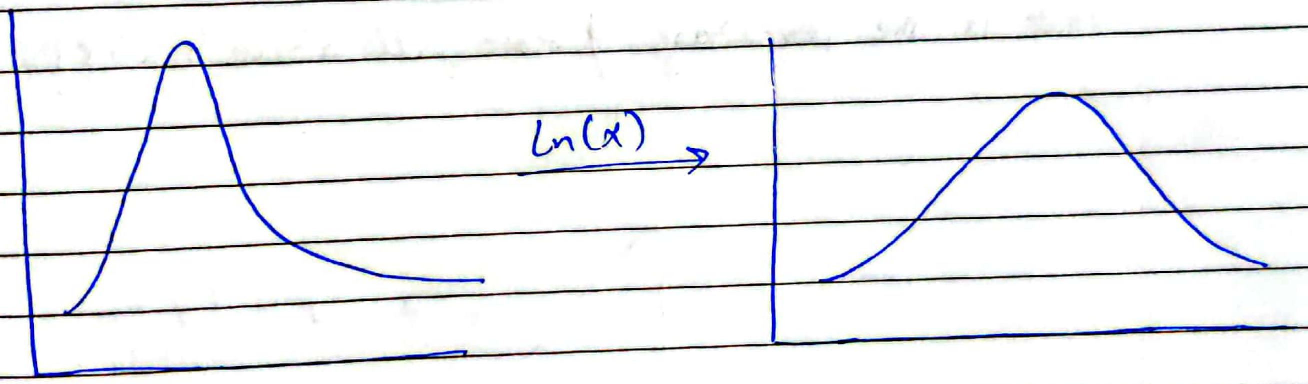
Log normal Dist is for Right Skewed Data Only

Let X be a random variable following Log normal distribution.

$X \sim \text{Log normal}(\mu, \sigma)$

↓

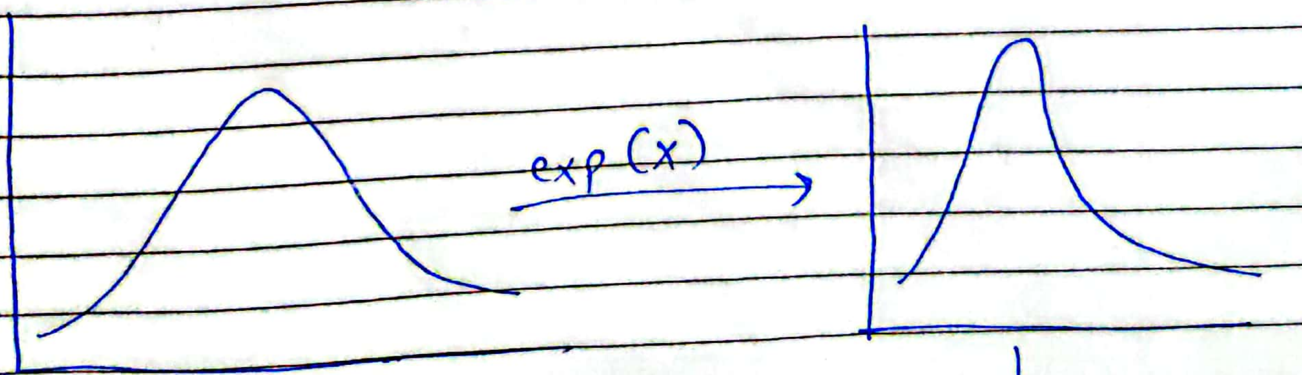
If X is log normally distributed then
$Y = \ln(X)$ have normal distribution



$X \sim$ Log normal

$y = \ln(X)$

now



$X \sim$ Normal Dist (μ, σ)

↓

Log - normal Distribution.

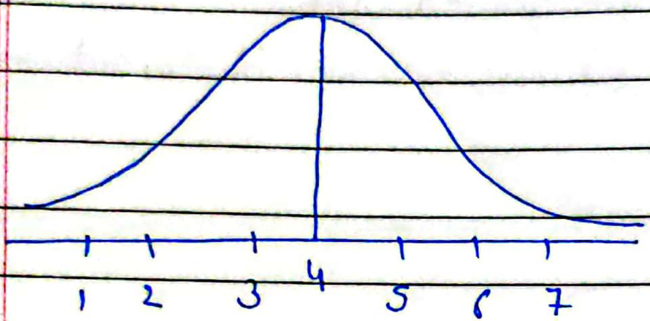
Q $X = [1, 2, 3, 4, 5, 6, 7]$

$\downarrow$

~~Let $\mu = 4$~~

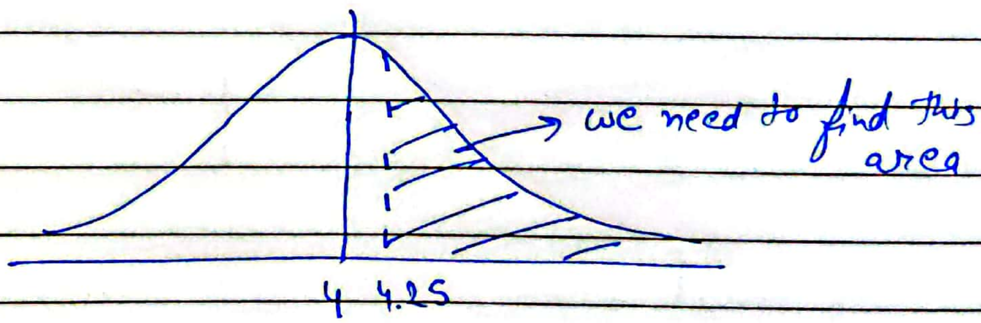
Let $\mu = 4$ and $\sigma = 1$

and it follows Normal Distribution



What is the percentage of score falls above 4.25?

Sol :-



① $Z\text{-score} = \dfrac{x-\mu}{\sigma} = \dfrac{4.25-4}{1} = \boxed{0.25}$ (So it is 0.25 std. dev away from μ)

Z-table (area under the curve)

↓

From Z-table $0.25 \rightarrow 0.5987$

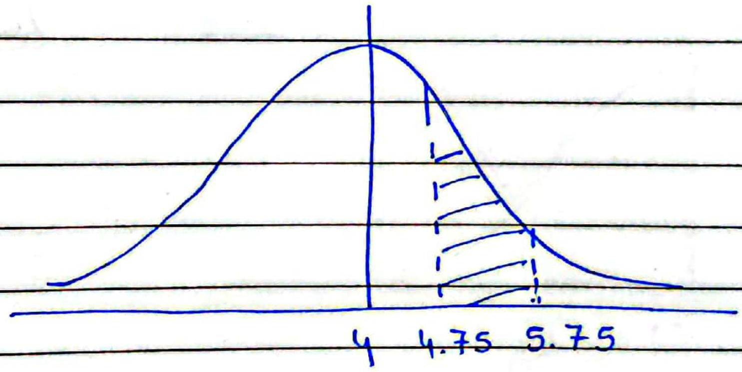
↓

Required area $= 1 - 0.5987$

or

$(100 - 59)\% \rightarrow 41\%$ approx

a % of score falls b/w 4.75 and 5.75?



$$Z\text{-score}(4.75) = \frac{4.75 - 4}{1} = 0.75 = 0.77337$$

$$Z\text{-score}(5.75) = \frac{5.75 - 4}{1} = 1.75 = 0.95994$$

$= 0.95994 - 0.77337$

$= 0.18657$

$= 18.65\%$