① Standard Normal Distribution

Z-score

which is done using the

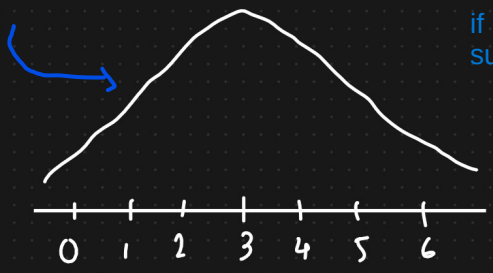
it is called as standard normal distribution

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

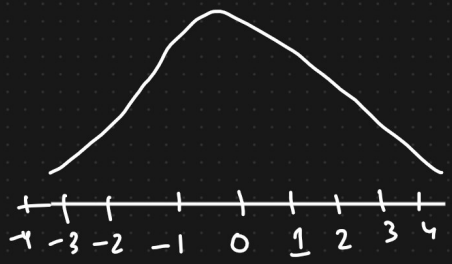
$\mu = 3$

$\sigma = 1.414 \approx 1$

this is my normal distribution



if i transform it such that $\Longrightarrow$ $\mu = 0$ and $\sigma = 1$



$X = \{1, 2, 3, 4, 5\}$ $\quad$ Z-score $= \dfrac{x_i - \mu}{\sigma}$ $\quad$ $Y = \{-2, -1, 0, 1, 2\}$

required std distribution

① $\dfrac{1-3}{1} = -2$ $\quad$ ③ $\dfrac{3-3}{1} = 0$

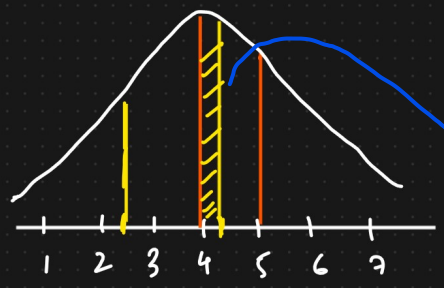
② $\dfrac{2-3}{1} = -1$ $\quad$ ④ $\dfrac{4-3}{1} = 1$

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

a random variable belongs to a standard normal distribution if mean =1 and std deviation =1

random variable

* imp interview question :



$\mu = 4$

$\sigma = 1$

4.25 will fall between 4 and 5

Q) How many standard deviation 4.25 is away from the mean ?

$x_i = 4.25$

$$z\text{-score} = \frac{4.25 - 4}{1} = 0.25$$

this indicates that 4.25 is 0.25 away from the mean

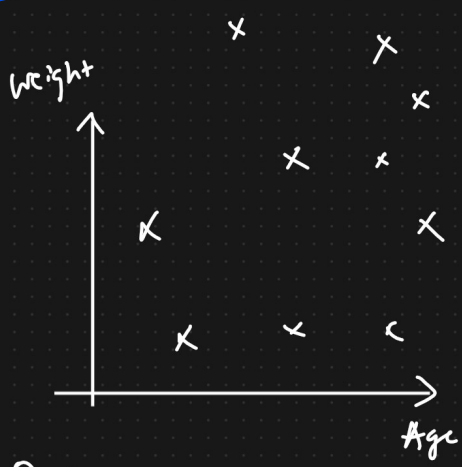
$x_i = 2.5$

$$z\text{-score} = \frac{2.5 - 4}{1} = -1.5$$

why are we using the z score :

Eg : Dataset (consider example of a data set)

Age (years)	Weight (kg)	Height (cms)	Salary (INR)
24	70	175	40K
25	60	160	50K
25	55	180	60K
22	40	130	30K
30	30	175	20K
31	25	180	70K
⇓	⇓	⇓	⇓



① Clustering Algorithms
② Linear Regression
③ Logistic Regression

this can be done using ML algorithm like :

if i try to plot the graph with such dataset my datapoints would be scattered like this as every value has different units :

Standardization ⇒ ML Models

for that we apply a feature called as :

$$z\text{-score} = \frac{x_i - \mu_{age}}{\sigma} \qquad \frac{x_i - \mu_{weight}}{\sigma}$$

here we apply the z score formula for every unit (like age , weight , height etc)

by this we are able to bring performance and effiiciecy to the model. This is called as standardization.