Lec-II

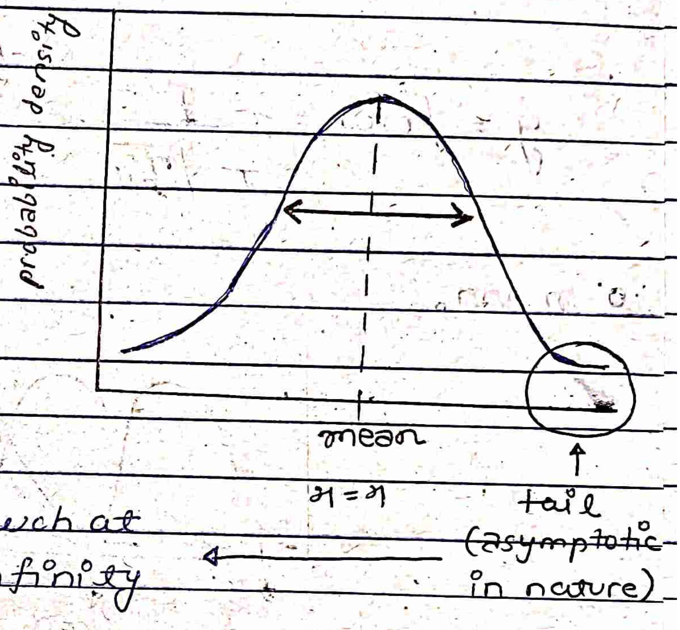
NORMAL DISTRIBUTION

★ **What is Normal Distribution?**

- Normal distribution also known as Gaussian distribution, is a propability distribution that is commonly used in statistical analysis. It's a continuous distribution that's symmetrical around the mean, with the Bell shaped curve.

properties:

→ Taile
→ Asymptotic in nature
→ Lots of points near the mean and very few far away
→ symmetric about mean



- The normal distribution is characterized by two parameters: the mean (μ) and the standard deviation (σ). The mean represents the centre of the distribution, while the standard deviation represents the spread of distribution.

$$X \sim N(\mu, \sigma^2)$$

X is normal random variable
σ standard deviation, $\sigma \geq 0$
μ mean, $-\infty \leq \mu \leq \infty$

Why is it so important?

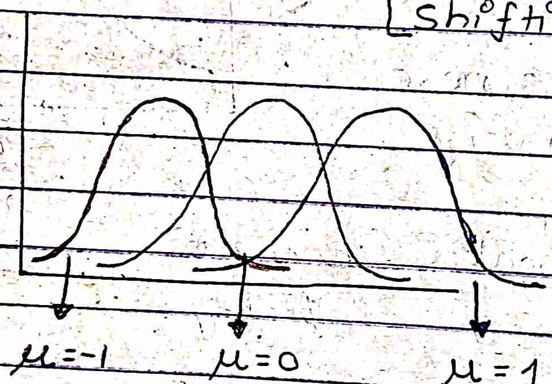
- Commonly in nature many natural phenomena follows a normal distribution, such as the heights of people, the weights of Objects, the IQ score of a population and many more.

Thus, the Normal distribution provides a convinient way to model and Analyze such data.

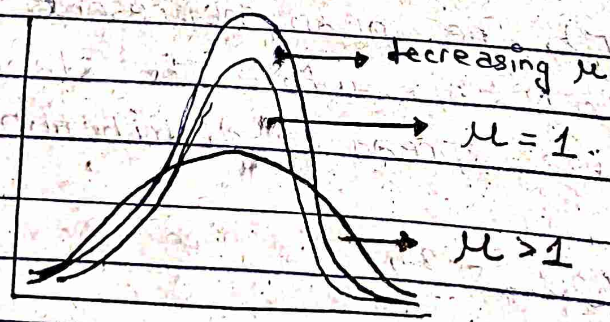
PDF Equation of Normal Distribution

$$y = f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

- mean, [shifting on x-axis]

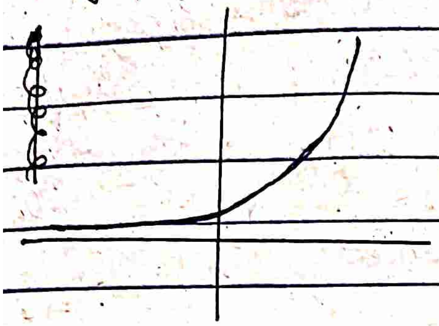


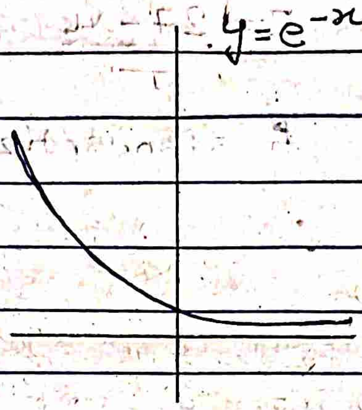
- standard deviation, [skewness]

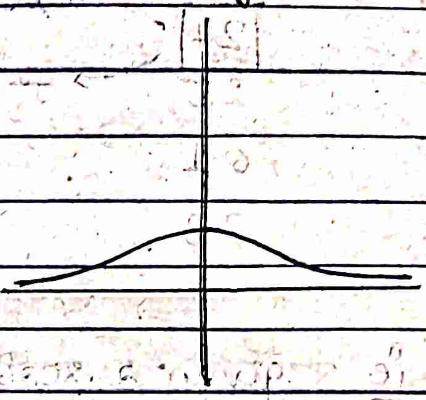


○ equation and it's parameters :

$y = e^{x}$ $\quad$ $y = e^{-x}$ $\quad$ $y = e^{-x^2}$







exponential growth $\quad$ exponential decay

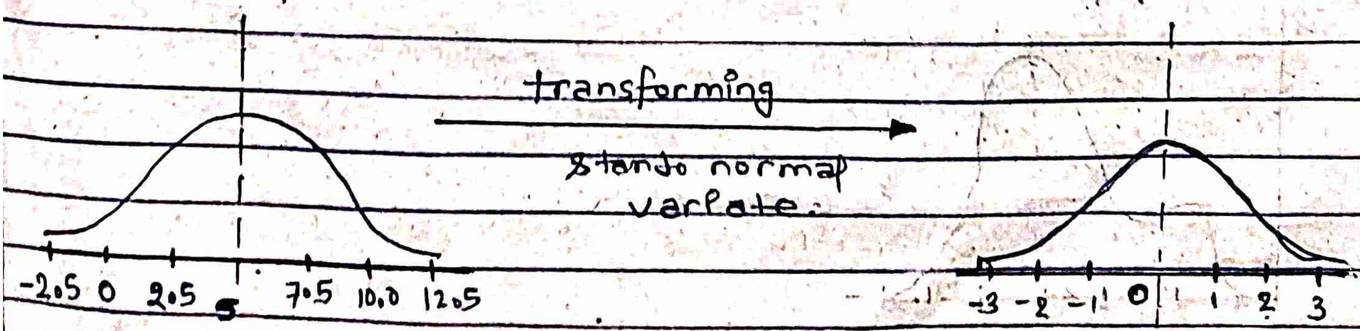
$$y = e^{-\frac{(x-\mu)^2}{2\sigma^2}} \rightarrow \text{Accomodate various curves}$$

★ Standard Normal Variate (Z) : A standard normal variate is a standard form of Normal distribution with $\boxed{\text{mean} = 0}$ and $\boxed{\text{standard deviation} = 1}$

Standardizing a normal distribution allows us to compare different distributions with each other, and to calculate probabilities using standardized tables or software.

$$\Rightarrow \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

iQ $X \sim N(5, 2.5)$



o. transformation,

ie Age

$\boxed{27}$ → μ, → σ

25

61

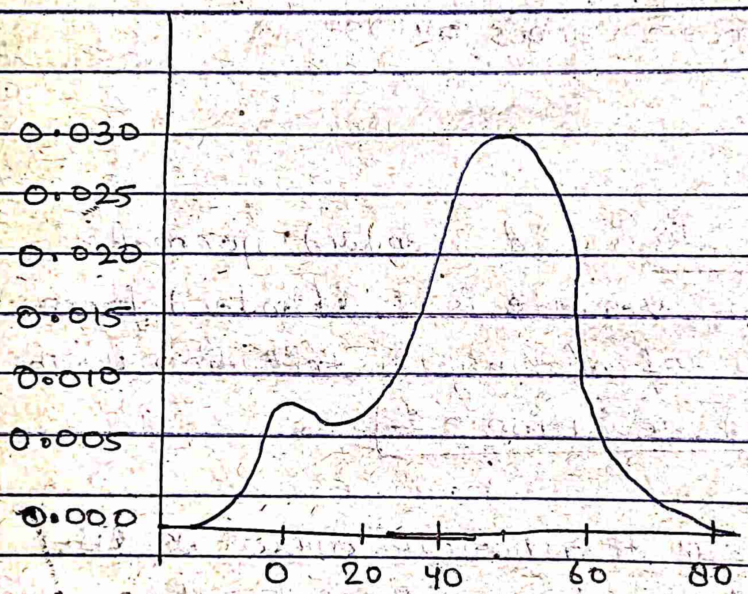
73

$z = \frac{27-\mu}{\sigma}$

to standardize

ie given a seaborn titanic dista

sns. kdeplot (titanic ['Age']



titanic ['Age']. mean()
o/p → 29.69911

titanic ['Age'].std()
o/p → 14.5264

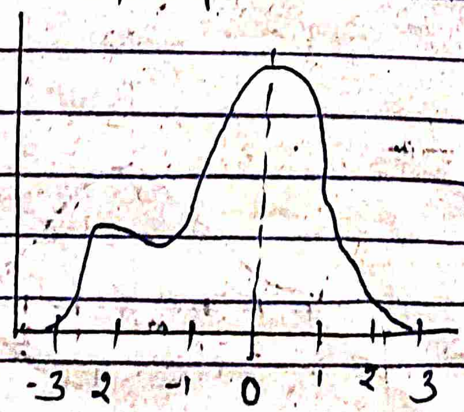
⇑

N (29, 14)

N(29, 14) $\xrightarrow{\text{transform}}$ N(0,1)

x = (titanic ['Age'] - titanic ['Age']. mean()) / titanic ['Age'].std()

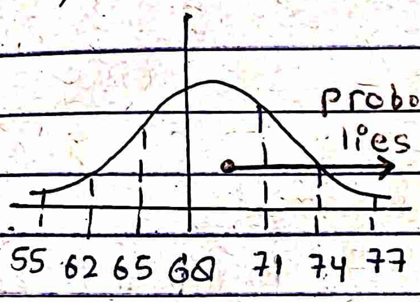
sns. kdeplot (x)



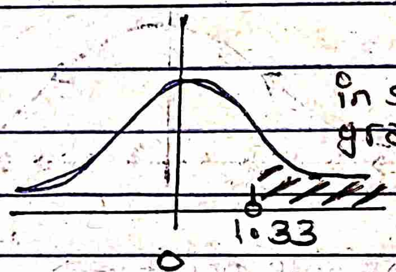
Q) Suppose the height of certain males in a certain population follows a normal distribution with a mean of 68 inches and a standard deviation of 3 inches; what is the probability that a randomly selected adult male from this population is taller than 72 inches?

$X \sim N(68, 3)$

$$Z = \frac{72-68}{3} \Rightarrow 4/3 \text{ or } 1.33$$



probo. of being taller lies in this region.

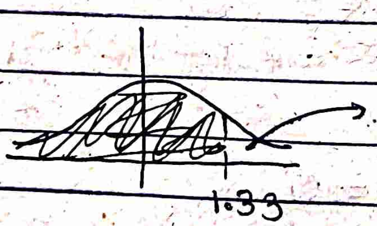


in standard graph.

○ with help of Z-tables →

- A table that tells you the area underneath a normal distribution curve, to the left of the Z-score.

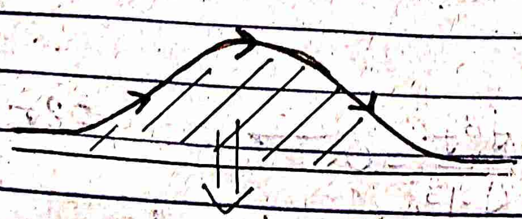
- population is 72 inches or lesser, probo. for 1.33 Z score on Z table → 0.90824 or 90.8 %.



100 % − 90.8 %

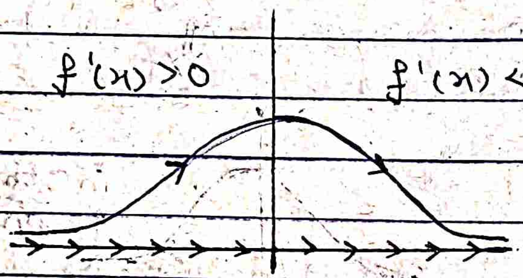
⇒ 9 % greater than 72 inches.

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = 1$$



total Area = 1

Empirical Rule: The normal distribution has well known empirical-rule also called the 68–95–99.7 rule, which states that approximately 68% of the data falls within one standard deviation of the mean about 95% of data falls within two standard deviati of the mean and about 99.7% of data falls within three standard deviation of the mean.



infinite point → to find probability at any given point.

$\frac{d^2 f}{dx^2} = 0$ (point of inflexion)

(sudden change in curve)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

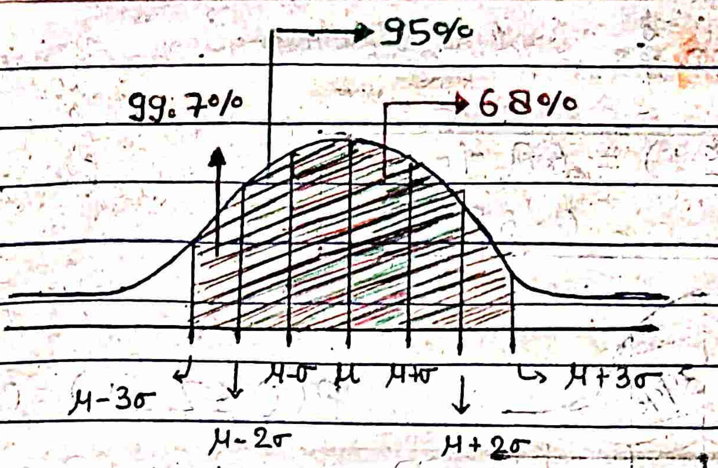
$\downarrow \frac{d^2 f}{dx^2}$

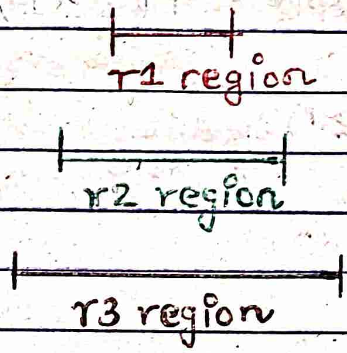
$$\frac{-1}{\sigma^3\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \left[1 - \frac{(x-\mu)^2}{\sigma^2}\right] = 0$$

$\frac{df}{dx} = 0 \longrightarrow x = \mu$

$\frac{d^3 f}{dx^3} = 0 \longrightarrow x = \mu \pm 2\sigma$

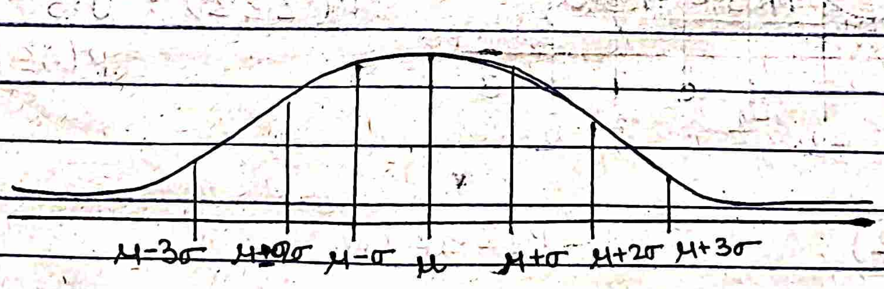
$\frac{d^2 f}{dx^2} = 0 \longrightarrow x = \mu \pm \sigma$





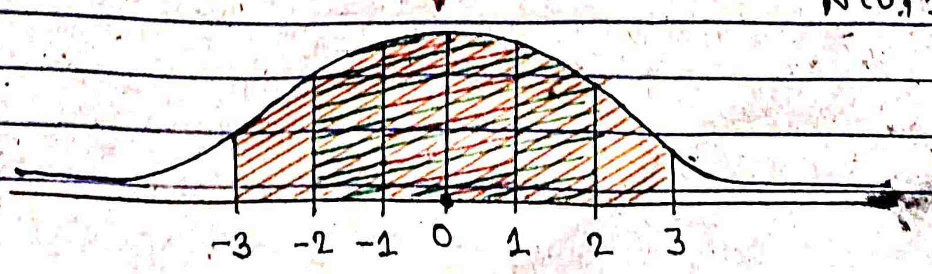
Shifting the origin :

normal distribution (μ, σ) $\xrightarrow[\text{change \{transform\}}]{\mu = 0,\ \sigma = 1}$ standard normal random variable.



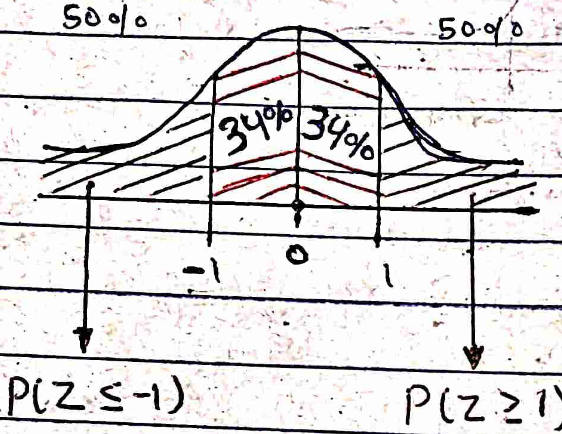
$\mu = 0, \sigma = 1$ (transform)

N(0, 1)



→ **r1 region:**

$$\boxed{P(-1 \le Z \le 1) = 68\% \text{ or } 0.6834}$$

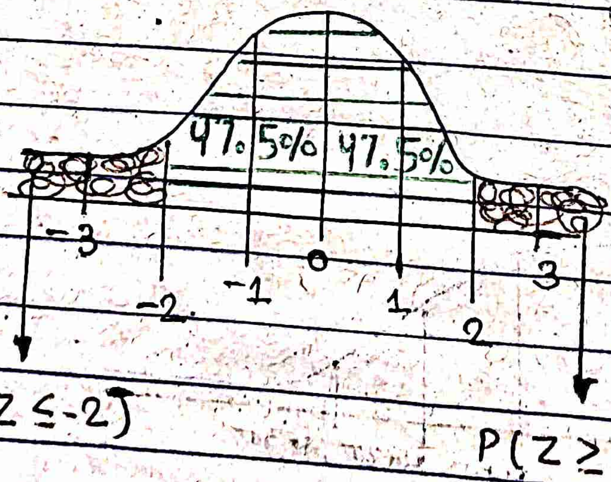


$P(0 \le Z \le 1) = 0.3147$

$P(Z \ge 1) = 0.5 - P(0 \le Z \le 1)$

→ **r2 region:**

$$\boxed{P(-2 \le Z \le 2) = 95\% \text{ or } 0.9547}$$



$P(0 \le Z \le 2) = 0.4773$

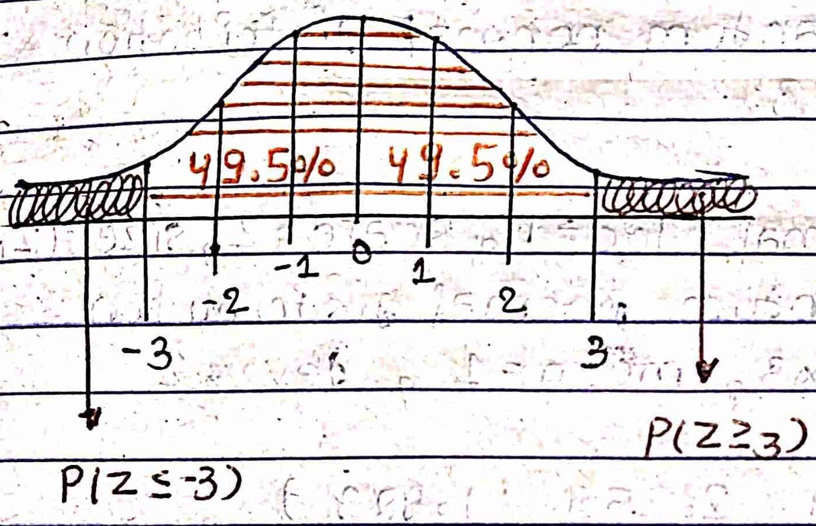
$P(Z \ge 2) = 0.5 - P(0 \le Z \le 2)$

→ **3 region :**

$P(-3 \le Z \le 3) = 0.9971$ or 99.7%

$P(0 \le Z \le 3) = 0.498$

$P(Z \ge 3) = 0.5 - P(0 \le Z \le 3)$



Visualization : It fits the probability distribution of many events.

ie IQ scores, heartbeats etc.

Use the random.normal() method to get a normal data distribution.

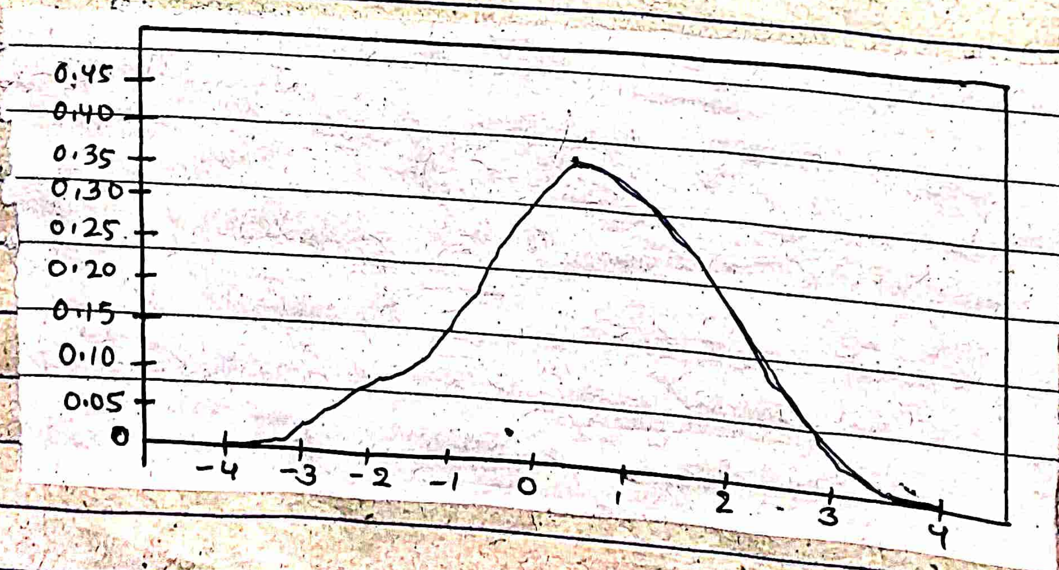
It has three parameters —

- loc – (mean) where the peak of bell exists
- scale – (standard devo) how flat the graph distribution should be
- size – The shape of returned array.

```
from numpy import random
import seaborn as sns

x = random.normal(size=(2,3))
# generate a random normal distribution
    of size 2x3.

x = random.normal(loc=1, scale=2, size=(2,
# generate a random normal distribution
# with size 2x3, mean=1, devo=2
```

O/P → [[0.321 2.259 1.983]
 [1.575 2.054 -0.622]]

```
sns.distplot(random.normal(size=1000),
                          hist=False)

plt.show()
```



$P(a \le X \le b)$

normal random variable

$N(\mu, \sigma^2)$

$$P\left[\frac{a-\mu}{\sigma} \le \overset{Z}{\frac{X-\mu}{\sigma}} \le \frac{b-\mu}{\sigma}\right]$$

−1 or 1   −2 or 2   −3 or 3

$P(X \ge a)$

z-score

$$P\left(\frac{x-\mu}{\sigma} \ge \frac{a-\mu}{\sigma}\right)$$

$$P\left(Z \ge \frac{a-\mu}{\sigma}\right)$$

Q) A nationalized bank has found that the daily balance available in it's savings account follows a normal distribution with a mean of ₹500 & a std. devi of ₹50. The percentage of savings account holder, who maintains an average daily balance more than ₹500 is—

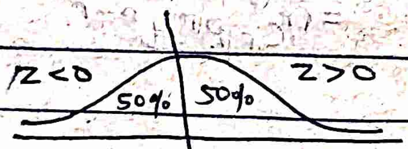
$\mu = ₹500$, $\sigma = ₹50$

$P(X > 500)$

normal random variable

$$P\left(\frac{X-\mu}{\sigma} > \frac{500-\mu}{\sigma}\right)$$

$$P\left(Z > \frac{500-500}{50}\right)$$



$\Rightarrow P(Z>0) \Rightarrow 0.5$ or 50%

लोग ऐसे है, जिनके bank मे ₹500 से ज्यादा है।