

A solid red vertical bar is positioned on the left side of the slide. To its right, there is a small blue circle with a white outline.

Module-3

Part – 6

Normal distribution

Normal Distribution

The normal distribution is a continuous distribution, is the most important of all the distributions.

It can be derived from the binomial distribution in the limiting case when n is very large and p is close to 0.5.

Normal Distribution

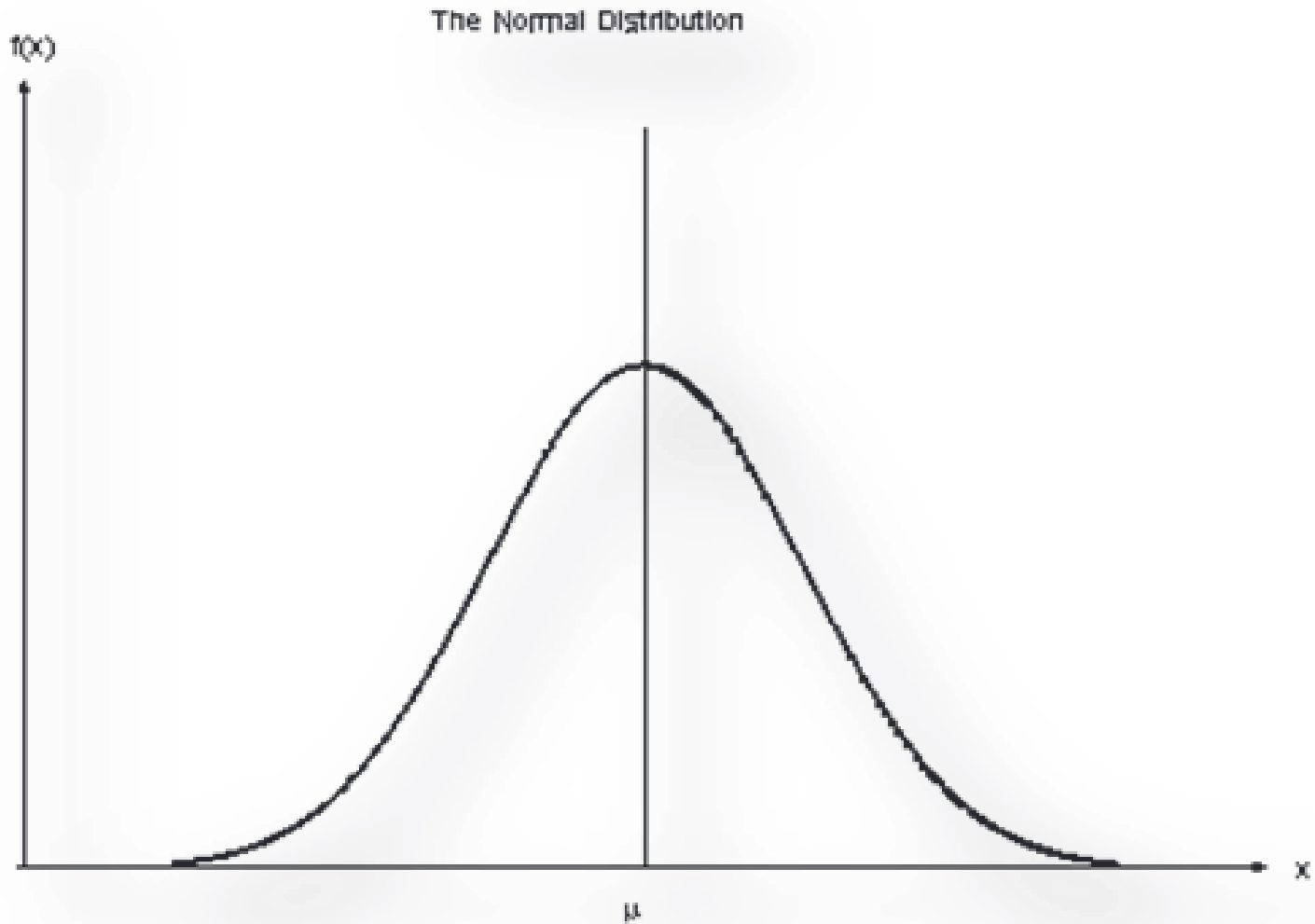
The normal distribution has two parameters (two numerical descriptive measures): the mean (μ) and the standard deviation (σ). If X is a quantity to be measured that has a normal distribution with mean (μ) and standard deviation (σ), we designate this by writing the following formula of the normal probability density function:


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

where $-\infty < x < \infty$, $-\infty < \mu < \infty$ and $\sigma > 0$.

Here x is called normal variate and $f(x)$ is called probability density function of the normal distribution.

Graph of the Normal distribution



- 
- The graph of the normal distribution is called the **normal curve**.
 - It is bell-shaped and symmetrical about the line

$$x = \mu$$

- The line $x = \mu$ divides the total area under the curve which is equal to 1 into two equal parts.
- The area to the right as well as to the left of the line $x = \mu$ is 0.5.

Mean and variance of Normal distribution

The mean and variance of the normal distribution are given by

$$\text{Mean} = \mu$$

$$\text{Variance } V = \sigma^2$$

$$\text{Standard deviation} = \sigma$$

Standard form of the Normal distribution

If X is a normal random variable with mean μ and standard deviation σ then the random variable $z = \frac{x - \mu}{\sigma}$ has the normal distribution with mean 0 and standard deviation 1. The random variable z is called the standard normal random variable.

The standard normal distribution is a normal distribution of standardized values called **z-scores**.

A z-score is measured in units of the standard deviation. The mean for the standard normal distribution is zero, and the standard deviation is one.

If X is a normally distributed random variable and $X \sim N(\mu, \sigma)$, then the z-score for a particular x is:

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

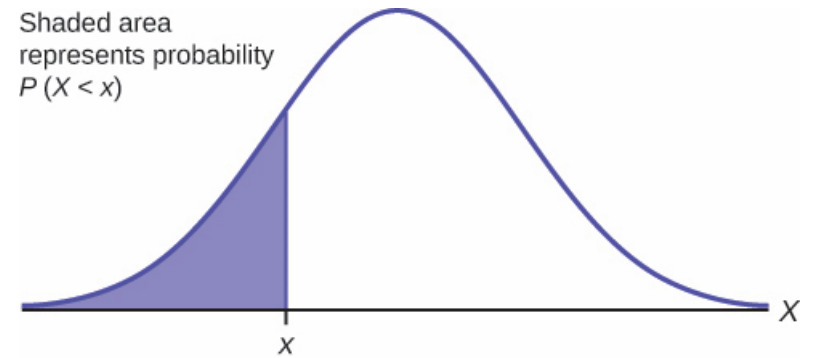
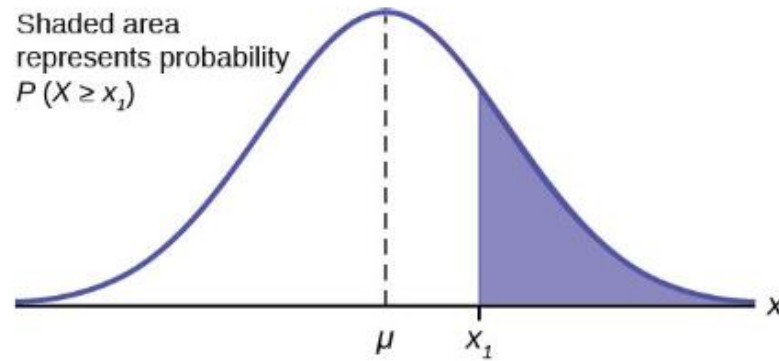
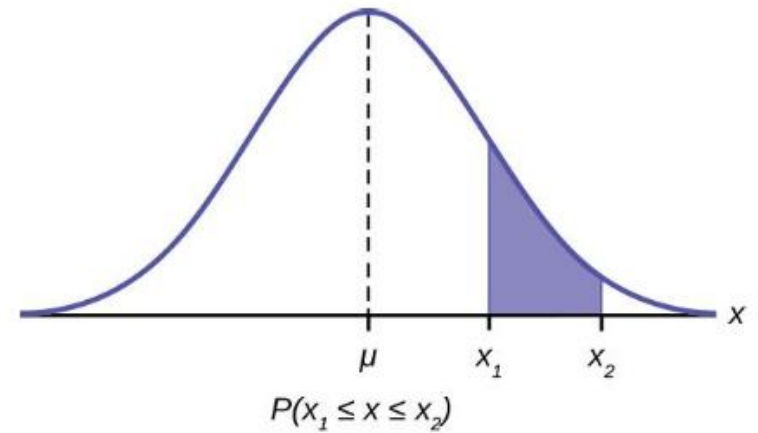
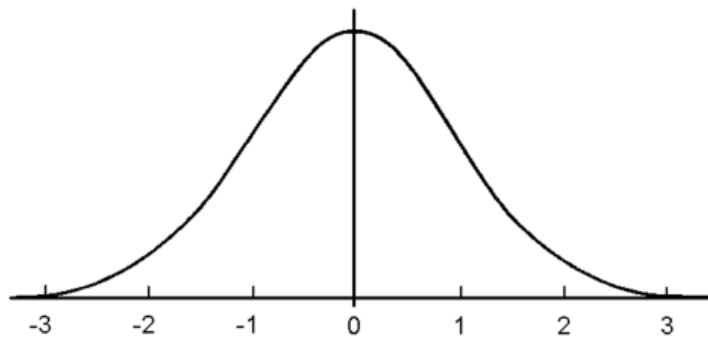
- The z-score tells you how many standard deviations the value x is above (to the right of) or below (to the left of) the mean, μ .
- Values of x that are larger than the mean have positive z-scores, and values of x that are smaller than the mean have negative z-scores.
- If x equals the mean, then x has a z-score of zero.

The probability density function for the normal distribution in standard form is given by

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

This helps us to compute areas under the normal probability curve by making use of standard tables.

The mathematical tool needed to find the area under a curve is integral calculus. The integral of the normal probability density function between the two points x_1 and x_2 is the area under the curve between these two points and is the probability between these two points.



Normal probability distribution curves

Note:

(1) If $f(z)$ is the probability density function for the normal distribution,

$$\text{then } P(z_1 < z < z_2) = \int_{z_1}^{z_2} f(z) dz = f(z_2) - f(z_1) \text{ where } f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

The function $f(z)$ defined above is called the distribution function for the normal distribution.

(2) The probabilities $P(z_1 \leq z \leq z_2)$, $P(z_1 \leq z < z_2)$, $P(z_1 < z \leq z_2)$ and $P(z_1 < z < z_2)$ are all regarded to be the same.

$$(3) \quad P(z \leq z_1) = \int_{-\infty}^{z_1} f(z) dz \text{ and } P(z \geq z_1) = \int_{z_1}^{\infty} f(z) dz$$

$$(4) \quad P(-\infty \leq z \leq 0) = P(0 \leq z \leq \infty) = 0.5$$

$$(5) \quad P(z \geq z_1) = P(0 \leq z \leq \infty) - P(0 \leq z \leq z_1)$$

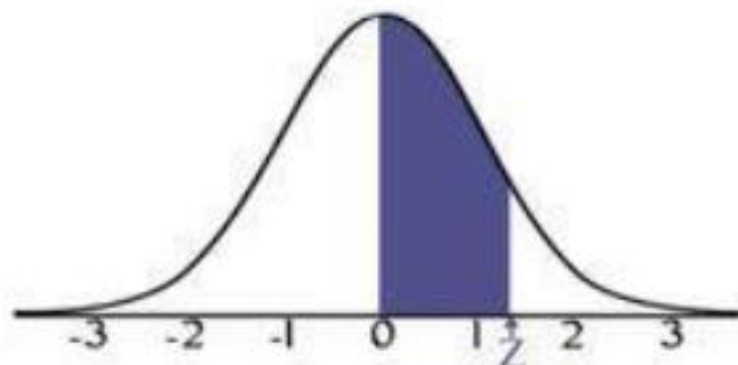
$$(6) \quad P(-z_1 \leq z \leq z_2) = P(-z_1 \leq z \leq 0) + P(0 \leq z \leq z_2)$$

$$(7) \quad P(-z_1 < z < 0) = P(0 < z < z_1) \text{ (since the normal curve is symmetric)}$$

$$(8) \quad P(-z_1 < z < z_1) = 2P(0 < z < z_1)$$

$$(9) \quad F(-z_1) = 1 - F(z_1)$$

- (10) The following table called normal probability table gives the area under the standard normal curve from 0 to $z > 0$, for various values of z .




STANDARD NORMAL TABLE (Z)

Entries in the table give the area under the curve between the mean and z standard deviations above the mean. For example, for $z = 1.25$ the area under the curve between the mean (0) and z is 0.3944.

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0190	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2969	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3513	0.3554	0.3577	0.3529	0.3621

[illegible]

- 
- It is also important to note that because the normal distribution is symmetrical it does not matter if the z-score is positive or negative when calculating a probability.
 - One standard deviation to the left (negative z-score) covers the same area as one standard deviation to the right (positive z-score).
 - This fact is why the Standard Normal tables do not provide areas for the left side of the distribution.

Example If X is a normal variate with mean 30 and S.D. 5, find the probabilities that (i) $26 \leq X \leq 40$, (ii) $X \geq 45$ and (iii) $|X - 30| > 5$.

Given, $\mu = 30$ and S.D. $\sigma = 5$

We have,
$$z = \frac{x - \mu}{\sigma}$$

$$\Rightarrow z = \frac{x - 30}{5}$$

(i) When $x = 26 \Rightarrow z = \frac{26 - 30}{5} = -0.8$

When $x = 40 \Rightarrow z = \frac{40 - 30}{5} = 2$

$$\therefore P(26 \leq X \leq 40) = P(-0.8 \leq z \leq 2)$$

$$\begin{aligned}
 \therefore P(26 \leq X \leq 40) &= P(-0.8 \leq z \leq 2) \\
 &= P(-0.8 \leq z \leq 0) + P(0 \leq z \leq 2) \\
 &= P(0 \leq z \leq 0.8) + P(0 \leq z \leq 2) \\
 &= 0.2881 + 0.4772 = 0.7653
 \end{aligned}$$

(ii) When $x = 45 \Rightarrow z = \frac{45 - 30}{5} = 3$

$$\begin{aligned}
 \therefore P(x \geq 45) &= P(z \geq 3) \\
 &= P(0 \leq z \leq \infty) - P(0 \leq z \leq 3) \\
 &= 0.5 - 0.4986 = 0.0014
 \end{aligned}$$

(iii) $P[|X - 30| \leq 5] = P(25 \leq X \leq 35)$

When $x = 25 \Rightarrow z = \frac{25 - 30}{5} = -1$

When $x = 35 \Rightarrow z = \frac{35 - 30}{5} = 1$

$$\begin{aligned}\therefore P(25 \leq X \leq 35) &= P(-1 \leq z \leq 1) \\ &= 2P(0 \leq z \leq 1) \\ &= 2(0.3413) = 0.6826\end{aligned}$$

$$\begin{aligned}\therefore P[|X - 30| > 5] &= 1 - P[|X - 30| \leq 5] \\ &= 1 - 0.6826 = 0.3174\end{aligned}$$

Example In a test on 2000 electric bulbs, it was found that the life of a particular make was normally distributed with an average life of 2040 hours and S.D. of 60 hours. Estimate the number of bulbs likely to burn for (i) more than 2150 hours, (ii) less than 1950 hours, (iii) more than 1920 hours and less than 2160 hours.

Given, $\mu = 2040$ hours and S.D. $\sigma = 60$ hours

We have,
$$z = \frac{x - \mu}{\sigma}$$

$$\Rightarrow z = \frac{x - 2040}{60}$$

(i) When $x = 2150 \Rightarrow z = \frac{2150 - 2040}{60} = 1.83$

$$\therefore P(x > 2150) = P(z > 1.83)$$

$$\begin{aligned}
 \therefore P(x > 2150) &= P(z > 1.83) \\
 &= P(0 < z < \infty) - P(0 < z < 1.83) \\
 &= 0.5 - 0.4664 = 0.0336
 \end{aligned}$$

\therefore Number of bulbs expected to burn for more than 2150 hours is
 $= 0.0336 \times 2000 \approx 67$

(ii) When $x = 1950 \Rightarrow z = \frac{1950 - 2040}{60} = -1.5$

$$\begin{aligned}
 \therefore P(x < 1950) &= P(z < -1.5) \\
 &= P(z > 1.5) \\
 &= P(0 < z < \infty) - P(0 < z < 1.5)
 \end{aligned}$$

(ii) When $x = 1950 \Rightarrow z = \frac{1950 - 2040}{60} = -1.5$

$$\therefore P(x < 1950) = P(z < -1.5)$$


$$= P(z > 1.5)$$

$$= P(0 < z < \infty) - P(0 < z < 1.5)$$

$$= 0.5 - 0.4332 = 0.0668$$

\therefore Number of bulbs expected to burn for more than 2150 hours is

$$= 0.0668 \times 2000 \approx 134$$



(iii) When $x = 1920 \Rightarrow z = \frac{1920 - 2040}{60} = -2$

When $x = 2160 \Rightarrow z = \frac{2160 - 2040}{60} = 2$

$$\begin{aligned}\therefore P(1920 < x < 2160) &= P(-2 < z < 2) \\ &= 2P(0 < z < 2) \\ &= 2(0.4772) = 0.9544\end{aligned}$$

\therefore Number of bulbs expected to burn for more than 1920 hours and less than 2160 hours is

$$= 0.9544 \times 2000 \approx 1909$$

Example A sample of 100 dry battery cells tested to find the length of life produced the following results:

$$\bar{x} = 12 \text{ hours and S.D. } \sigma = 3 \text{ hours}$$

Assuming the data to be normally distributed, what percentage of battery cells are expected to have life (i) more than 15 hours, (ii) less than 6 hours and (iii) between 10 and 14 hours?

Given, $\mu = 12$ hours and S.D. $\sigma = 3$ hours

$$\text{We have, } z = \frac{x - \mu}{\sigma}$$

$$\Rightarrow z = \frac{x - 12}{3}$$

$$(i) \quad \text{When } x = 15 \Rightarrow z = \frac{15 - 12}{3} = 1$$

$$\therefore P(x > 15) = P(z > 1)$$

$$\begin{aligned}
 \therefore P(x > 15) &= P(z > 1) \\
 &= P(0 < z < \infty) - P(0 < z < 1) \\
 &= 0.5 - 0.3413 = 0.1587 = 15.87\%
 \end{aligned}$$

(ii) When $x = 6 \Rightarrow z = \frac{6 - 12}{3} = -2$

$$\begin{aligned}
 \therefore P(x < 6) &= P(z < -2) \\
 &= P(z > 2) \\
 &= P(0 < z < \infty) - P(0 < z < 2) \\
 &= 0.5 - 0.4772 = 0.0228 = 2.28\%
 \end{aligned}$$

(iii) When $x = 10 \Rightarrow z = \frac{10 - 12}{3} = \frac{-2}{3} = -0.67$

When $x = 14 \Rightarrow z = \frac{14 - 12}{3} = \frac{2}{3} = 0.67$

$$\begin{aligned}\therefore P(10 < x < 14) &= P(-0.67 < z < 0.67) \\ &= 2P(0 < z < 0.67) \\ &= 2(0.2486) = 0.4972 = 49.72\%\end{aligned}$$

Example The marks of 1000 students in an examination follow a normal distribution with mean 70 and standard deviation 5. Find the number of students whose marks will be (i) less than 65, (ii) more than 75, (iii) between 65 and 75.

Given, $\mu = 70$ and S.D. $\sigma = 5$

We have,
$$z = \frac{x - \mu}{\sigma}$$

$$\Rightarrow z = \frac{x - 70}{5}$$

(i) When $x = 65 \Rightarrow z = \frac{65 - 70}{5} = -1$

$$\therefore P(x < 65) = P(z < -1)$$


$$\begin{aligned}
 \therefore P(x < 65) &= P(z < -1) \\
 &= P(z > 1) \\
 &= P(0 < z < \infty) - P(0 < z < 1) \\
 &= 0.5 - 0.3413 = 0.1587
 \end{aligned}$$

\therefore Number of students scoring less than 65 marks is

$$= 0.1587 \times 1000 \approx 159$$

(ii) When $x = 75 \Rightarrow z = \frac{75 - 70}{5} = 1$

$$\therefore P(x > 75) = P(z > 1)$$



$$\begin{aligned}\therefore P(x > 75) &= P(z > 1) \\ &= P(0 < z < \infty) - P(0 < z < 1) \\ &= 0.5 - 0.3413 = 0.1587\end{aligned}$$

\therefore Number of students scoring more than 75 marks is
 $= 0.1587 \times 1000 \approx 159$

(iii) When $x = 65 \Rightarrow z = -1$

When $x = 75 \Rightarrow z = 1$

$$\therefore P(65 < x < 75) = P(-1 < z < 1)$$


$$\begin{aligned}\therefore P(65 < x < 75) &= P(-1 < z < 1) \\ &= 2P(0 < z < 1) \\ &= 2(0.3413) = 0.6826\end{aligned}$$

\therefore Number of students scoring marks between 65 and 75 is
 $= 0.6826 \times 1000 \approx 683$

Example In an examination taken by 500 candidates, the average and the standard deviation of marks obtained (normally distributed) are 40% and 10% respectively. Find approximately (i) how many will pass, if 50% is fixed as a minimum, (ii) how many have scored marks above 60% and (iii) what should be the minimum percentage if 350 candidates are to pass?


Given, $\mu = 40$ and S.D. $\sigma = 10$

$$\text{We have, } z = \frac{x - \mu}{\sigma}$$

$$\Rightarrow z = \frac{x - 40}{10}$$

$$(i) \quad \text{When } x = 50 \Rightarrow z = \frac{50 - 40}{10} = 1$$

$$\therefore P(x \geq 50) = P(z > 1)$$


$$\begin{aligned}\therefore P(x \geq 50) &= P(z > 1) \\ &= P(0 < z < \infty) - P(0 < z < 1) \\ &= 0.5 - 0.3413 = 0.1587\end{aligned}$$

\therefore Number of students passing if 50% is fixed as minimum is

$$= 0.1587 \times 500 \approx 79$$

(ii) When $x = 60 \Rightarrow z = \frac{60 - 40}{10} = 2$

$$\therefore P(x > 60) = P(z > 2)$$

$$\begin{aligned}
 \therefore P(x > 60) &= P(z > 2) \\
 &= P(0 < z < \infty) - P(0 < z < 2) \\
 &= 0.5 - 0.4772 = 0.0228
 \end{aligned}$$

\therefore Number of students scoring more than 60% is

$$= 0.0228 \times 500 \approx 11$$

- (iii) Let M% of marks is the minimum for passing if 350 candidates are to pass. Then, we should have $M < 40$, and

$$P(x \geq M) \times 500 = 350$$

$$\Rightarrow P(x \geq M) = \frac{350}{500} = 0.7$$

We have, $z = \frac{x - 40}{10}$

$$\Rightarrow x = 10z + 40$$

$$\therefore P(x \geq M) = P(10z + 40 \geq M) = P\left(z \geq \frac{M - 40}{10}\right)$$

$$= P(z \geq -z_1) \text{ where } -z_1 = \frac{M - 40}{10}$$

$$= P(-z_1 \leq z \leq 0) + P(0 \leq z \leq \infty)$$

$$= P(0 \leq z \leq z_1) + 0.5$$

Since $P(x \geq M) = 0.7$

$$\Rightarrow P(0 \leq z \leq z_1) = 0.2$$

The normal probability table shows that $P(0 \leq z \leq z_1) = 0.2$ for $z_1 = 0.55$

$$\text{So, } -z_1 = \frac{M - 40}{10} = -0.55 \Rightarrow M = 34.5 \approx 35$$

\therefore 35% is fixed as minimum marks for passing in order to pass 350 candidates out of 500.

Example In a normal distribution, 31% of the items are under 45 and 8% are over 64. Find the mean and standard deviation of the distribution.

Let μ be the mean and σ be the S.D.

Given, 31% of the items are under 45

\Rightarrow Area to the left of the ordinate $x = 45$ is 0.31

Given, 8% of the items are over 64

\Rightarrow Area to the right of the ordinate $x = 64$ is 0.08

When $x = 45$, let $z = z_1$

$$P(z_1 < z < 0) = 0.5 - 0.31 = 0.19$$

From the normal probability table, the value of z corresponding to this area is 0.5.

$$\text{i.e., } z_1 = -0.5 \quad [\because z_1 < 0]$$

When $x = 64$, let $z = z_2$

$$P(0 < z < z_2) = 0.5 - 0.08 = 0.42$$

From the normal probability table, the value of z corresponding to this area is 1.4.

$$\text{i.e., } z_2 = 1.4$$

Since, $z = \frac{x - \mu}{\sigma}$

$$\Rightarrow -0.5 = \frac{45 - \mu}{\sigma} \text{ and } 1.4 = \frac{64 - \mu}{\sigma}$$

$$\Rightarrow \mu - 0.5\sigma = 45 \text{ and } \mu + 1.4\sigma = 64$$

Solving these equations, we get

$$\mu = 50 \text{ and } \sigma = 10$$

Example In a normal distribution, 7% of the items are under 35 and 89% are under 60. Find the mean and standard deviation of the distribution.

Let μ be the mean and σ be the S.D.

Given, 7% of the items are under 35

\Rightarrow Area to the left of the ordinate $x = 35$ is 0.07

Given, 89% of the items are under 60

\Rightarrow Area to the left of the ordinate $x = 60$ is 0.89

When $x = 35$, let $z = z_1$

$$P(z_1 < z < 0) = 0.5 - 0.07 = 0.43$$

From the normal probability table, the value of z corresponding to this area is 1.48

$$\text{i.e., } z_1 = -1.48 \quad [\because z_1 < 0]$$

When $x = 60$, let $z = z_2$

$$P(z_2 < z < 0) = 0.5 - 0.89 = -0.39$$

From the normal probability table, the value of z corresponding to this area is 1.23

$$\text{i.e., } z_2 = 1.23 \quad [\because z_2 > 0]$$

Since, $z = \frac{x - \mu}{\sigma}$

$$\Rightarrow -1.48 = \frac{35 - \mu}{\sigma} \text{ and } 1.23 = \frac{60 - \mu}{\sigma}$$

$$\Rightarrow \mu - 1.48\sigma = 35 \text{ and } \mu + 1.23\sigma = 60$$

Solving these equations, we get

$$\mu = 48.65 \text{ and } \sigma = 9.22$$

Example Fit a normal distribution for the data

x	2	4	6	8	10
f	1	4	6	4	1

$$\sum f_i = 1 + 4 + 6 + 4 + 1 = 16$$

$$\sum x_i f_i = 2 + 16 + 36 + 32 + 10 = 96$$

$$\sum x_i^2 f_i = 4 + 64 + 216 + 256 + 100 = 640$$

$$\therefore \text{Mean } \mu = \frac{\sum x_i f_i}{\sum f_i} = \frac{96}{16} = 6$$

$$\text{Variance } V = \frac{\sum x_i^2 f_i}{\sum f_i} - \mu^2 = \frac{640}{16} - 36 = 4$$

S.D $\sigma = \sqrt{V} = \sqrt{4} = 2$

The p.d.f. of normal distribution is

$$\begin{aligned} f(x) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \\ &= \frac{1}{2\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-6}{2}\right)^2} \end{aligned}$$

The equation of the normal curve that fits the data is

$$F(x_i) = \left(\sum f_i\right) f(x)$$

$$\Rightarrow F(x_i) = (16) \left[\frac{1}{2\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-6}{2}\right)^2} \right] = \frac{8}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-6}{2}\right)^2}$$

$$F(x_i) = (16) \left[\frac{1}{2\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-6}{2} \right)^2} \right] = \frac{8}{\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-6}{2} \right)^2}$$

$$F(2) = \frac{8}{\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{2-6}{2} \right)^2} = 0.97$$

$$F(4) = \frac{8}{\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{4-6}{2} \right)^2} = 3.9$$

$$F(6) = \frac{8}{\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{6-6}{2} \right)^2} = 6.1$$

$$F(8) = \frac{8}{\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{8-6}{2} \right)^2} = 3.9$$

$$F(10) = \frac{8}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{10-6}{2}\right)^2} = 0.97$$

\therefore The theoretical frequencies are

x	2	4	6	8	10
f	1	4	6	4	1