Normal Distribution

8·2. Normal Distribution. The normal distribution was first discovered in 1733 by English mathematician De-Moivre, who obtained this continuous distribution as a limiting case of the binomial distribution and applied it to problems arising in the game of chance. It was also known to Laplace, no later than 1774 but through a historical error it was credited to Gauss, who first made reference to it in the beginning of 19th century (1809), as the distribution of errors in Astronomy. Gauss used the normal curve to describe the theory of accidental errors of measurements involved in the calculation of orbits of heavenly bodies. Throughout the eighteenth and nineteeth centuries, various efforts were made to establish the normal model as the underlying law ruling all continuous random variables. Thus, the name "*normal*" . These efforts, however, failed because of false premises. The normal model has, nevertheless, become the most important probability model in statistical analysis.

Normal Distribution

A random variable X is said to have a normal distribution with parameters μ (called mean) and σ^2 (called variance) if its p.d.f is given by the probability law

$$f(x,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2\sigma^2}(x-\mu)^2}, \quad -\infty < x < \infty$$
$$-\infty < \mu < \infty, \sigma > 0$$

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

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

$$Z = \frac{X-\mu}{\sigma}$$, is a standard normal variate with

$E(Z) = 0$ and $Var(Z) = 1$

and we write $Z \sim N(0,1)$

3) The p.d.f of a standard normal variate Z is given by

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < \infty$$

and the corresponding distribution function is

$$\Phi(z) = P(Z \le z) = \int_{-\infty}^{z} \phi(u)\,du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2}\,du$$

Result (1) $\Phi(-z) = 1 - \Phi(z), \; z > 0$

§ Normal Distribution as a limiting case of Binomial distribution

Normal distribution is another limiting form of the binomial distribution under the following conditions:

(i) n, the number of trials is indefinitely large i.e. $n \to \infty$

(ii) neither p nor q is very small.

§ Chief Characteristic of the Normal distribution and Normal probability curve

The normal probability curve with mean μ and standard deviation σ is given by

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

and has the following properties

(1) The curve is bell-shaped and symmetrical about the line $x = \mu$

(II) Mean, median and mode of the distribution coincide.

(III) As x increases numerically, $f(x)$ decreases rapidly, the maximum probability occuring at the point $x = \mu$, and is given by $[p(x)]_{max} = \frac{1}{\sigma\sqrt{2\pi}}$

IV) x- axis is an asymptote to the curve.

The following table gives the area under the normal probability curve for some important values of the standard normal variate.

Distance from the mean ordinates in terms of $\pm\sigma$	Area under the curve
$z = \pm 0.745$	$50\% = 0.50$
$z = \pm 1.00$	$68.26\% = 0.6826$
$z = \pm 1.96$	$95\% = 0.95$
$z = \pm 2.0$	$95.44\% = 0.9544$
$z = \pm 2.58$	$99\% = 0.99$
$z = \pm 3.0$	$99.73\% = 0.9973$

§ Area under the normal distribution curve is 1.

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

let $u = \dfrac{x-\mu}{\sigma}$, $\sigma\, du = dx$.

$$I = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}u^2} \sigma\, du$$

$$I = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}u^2} du$$

$$I^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(u^2+v^2)} du\, dv$$

$u = r\cos\theta$, $v = r\sin\theta$

$$= \frac{1}{2\pi}\int_0^{2\pi}\int_0^{\infty} e^{-\frac{1}{2}r^2}\, r\,dr\,d\theta$$

$$= \frac{1}{2\pi}\int_0^{2\pi}\left[-e^{-\frac{1}{2}r^2}\right]_0^{\infty} d\theta$$

$$= \frac{1}{2\pi}\int_0^{2\pi} -\left[0-1\right] d\theta = \frac{1}{2\pi}\int_0^{2\pi} d\theta$$

$$= \frac{1}{2\pi} \times 2\pi = 1$$

§ Mode of the Normal distribution

Mode is the value of x for which $f(x)$ is maximum, i.e, mode is the solution of $f'(x) = 0$. and $f''(x) < 0$

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

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

$$\Rightarrow \frac{1}{f(x)} f'(x) = -\frac{1}{2\sigma^2} 2(x-\mu) = -\frac{1}{\sigma^2}(x-\mu)$$

$$\Rightarrow \quad f'(x) = -\frac{1}{\sigma^2}(x-\mu) f(x)$$

$$f'(x) = 0 \Rightarrow x = \mu$$

$$f''(x) = -\frac{1}{\sigma^2}\left[(x-\mu)f'(x) + f(x)\cdot 1\right]$$

$$= -\frac{1}{\sigma^2}\left[(x-\mu)\frac{(-1)}{\sigma^2}(x-\mu)f(x) + f(x)\right]$$

$$= \frac{-f(x)}{\sigma^2}\left[1 - \frac{(x-\mu)^2}{\sigma^2}\right]$$

At $x=\mu$, $f''(x) = -\dfrac{f(\mu)}{\sigma^2} = -\dfrac{1}{\sigma^2}\dfrac{1}{\sigma\sqrt{2\pi}} < 0$

So $f(x)$ is maximum at $x=\mu$

Hence $x=\mu$ is the mode of the normal distribution.

§ Median of Normal distribution

If M is the median of the normal distribution, we have

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

$$\Rightarrow \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\mu} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\,dx + \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu}^{M} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\,dx = 1/2$$

$$\Rightarrow \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu}^{M} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\,dx = 0$$

$$\left[\because \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\mu} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\,dx = 1/2\right]$$

$$\Rightarrow \underline{\mu = M}$$

Hence for normal distribution

Mean = Mode = Median.

MGF of Normal distribution

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx$$

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

let $z = \dfrac{x-\mu}{\sigma}$, $dz = \dfrac{1}{\sigma}\, dx$

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{t(\mu+\sigma z)}\, e^{-z^2/2}\, \sigma\, dz$$

$$= \frac{e^{t\mu}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z^2-2t\sigma z)}\, dz$$

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

$$= \frac{e^{t\mu}}{\sqrt{2\pi}} e^{\frac{t^2\sigma^2}{2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(z-\sigma t)^2} dz.$$

$$= e^{\mu t + \frac{t^2\sigma^2}{2}} \times \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}u^2} du$$

$$= e^{\mu t + \frac{1}{2}t^2\sigma^2} \times \frac{1}{\sqrt{2\pi}} \times \sqrt{2\pi} = e^{\mu t + \frac{1}{2}t^2\sigma^2}$$

$$M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

For Standard normal varcate

$$M_Z(t) = e^{\frac{1}{2}t^2}$$

§ Moments of Normal distribution

Odd order moments about mean are given by

$$\mu_{2n+1} = \int_{-\infty}^{\infty} (x-\mu)^{2n+1} f(x)\, dx$$

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

$$\mu_{2n+1} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z)^{2n+1} e^{-\frac{1}{2}z^2} dz$$

$$\left(z = \frac{x-\mu}{\sigma}\right)$$

$$= \frac{\sigma^{2n+1}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^{2n+1} e^{-\frac{1}{2}z^2} dz = 0$$

Since $z^{2n+1} e^{-\frac{1}{2}z^2}$ is an odd function

Even order moments about origin.

$$\mu_{2n} = \int_{-\infty}^{\infty} (x-\mu)^{2n} f(x)\,dx.$$

$$\Rightarrow \mu_{2n} = \sigma^2 (2n-1)\,\mu_{2n-2}$$ [Check the book for detail]

$$\boxed{\mu_{2n} = 1\cdot 3\cdot 5 \cdots (2n-1)\,\sigma^{2n}}$$

Example 8·12. *X is normally distributed and the mean of X is 12 and S.D. is 4. (a) Find out the probability of the following :*

(*i*) $X \geq 20$, (*ii*) $X \leq 20$, and (*iii*) $0 \leq X \leq 12$

(*b*) *Find* x', *when* $P(X > x') = 0{\cdot}24$.

(*c*) *Find* x_0' *and* x_1', *when* $P(x_0' < X < x_1') = 0{\cdot}50$ *and* $P(X > x_1') = 0{\cdot}25$

Solution. (*a*) We have $\mu = 12, \sigma = 4$, *i.e.*, $X \sim N(12, 16)$.

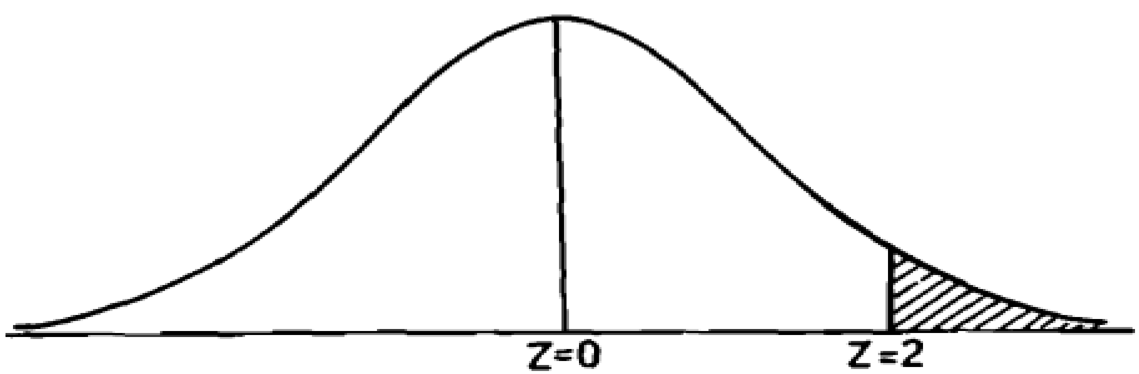
(*i*) $P(X \geq 20) = ?$

When $X = 20$, $\quad Z = \dfrac{20 - 12}{4} = 2$

$\therefore \quad P(X \geq 20) = P(Z \geq 2) = 0{\cdot}5 - P(0 \leq Z \leq 2) = 0{\cdot}5 - 0{\cdot}4772 = 0{\cdot}0228$

(*ii*) $P(X \leq 20) = 1 - P(X \geq 20)$ (∵ Total probability = 1)

$= 1 - 0{\cdot}0228 = 0{\cdot}9772$



(*iii*) $P(0 \leq X \leq 12) = P(-3 \leq Z \leq 0)$ $\left(Z = \dfrac{X - 12}{4}\right)$

$= P(0 \leq Z \leq 3) = 0{\cdot}49865$ (From symmetry)

Example 8·14. *The mean yield for one-acre plot is 662 kilos with a s.d. 32 kilos. Assuming normal distribution, how many one-acre plots in a batch of 1,000 plots would you expect to have yield (i) over 700 kilos, (ii) below 650 kilos, and (iii) what is the lowest yield of the best 100 plots?*

Solution. If the r.v. X denotes the yield (in kilos) for one-acre plot, then we are given that $X \sim N(\mu, \sigma^2)$, where $\mu = 662$ and $\sigma = 32$.

(*i*) The probability that a plot has a yield over 700 kilos is given by

$$P(X > 700) = P(Z > 1 \cdot 19)\,; \qquad Z = \frac{X - 662}{32}$$

$$= 0 \cdot 5 - P(0 \le Z \le 1 \cdot 19)$$

$$= 0 \cdot 5 - 0 \cdot 3830$$

$$= 0 \cdot 1170$$

Hence in a batch of 1.000 plots, the expected number of plots with yield over 700 kilos is $1{,}000 \times 0 \cdot 117 = 117$.

(*ii*) Required number of plots with yield below 650 kilos is given by

$$1000 \times P(X < 650) = 1000 \times P(Z < -0 \cdot 38) \qquad \left[Z = \frac{650 - 662}{32} \right]$$

$$= 1000 \times P(Z > 0 \cdot 38) \qquad \text{(By symmetry)}$$

$$= 1000 \times [\, 0 \cdot 5 - P(0 \le Z \le 0 \cdot 38) \,]$$

$$= 1000 \times [\, 0 \cdot 5 - 0 \cdot 1480 \,] = 1000 \times 0 \cdot 352$$

$$= 352$$

(*iii*) The lowest yield, say, x_1 of the best 100 plots is given by

$$P(X > x_1) = \frac{100}{1000} = 0{\cdot}1$$

When $$X = x_1, \ Z = \frac{x_1 - \mu}{\sigma} = \frac{x_1 - 662}{32} = z_1 \text{ (say)} \qquad \ldots(*)$$

such that $P(Z > z_1) = 0{\cdot}1 \quad \Rightarrow \quad P(0 \le Z \le z_1) = 0{\cdot}4$.

$\Rightarrow \qquad z_1 = 1{\cdot}28$ (approx.) [From Normal Probability Tables]

Substituting in (*), we get

$$x_1 = 662 + 32\, z = 662 + 32 \times 1{\cdot}28$$

$$= 662 + 40{\cdot}96 = 702{\cdot}96$$

Hence the best 100 plots have yield over 702·96 kilos.

Example 8·15. *There are six hundred Economics students in the post-graduate classes of a university, and the probability for any student to need a copy of a particular book from the university library on any day is* 0·05. *How many copies of the book should be kept in the university library so that the probability may be greater than 0·90 that none of the students needing a copy from the library has to come back disappointed? (Use normal approximation to the binomial distribution).*

Solution. We are given :

$$n = 600, p = 0{\cdot}05, \mu = np = 600 \times 0{\cdot}05 = 30$$

$$\sigma^2 = npq = 600 \times 0{\cdot}05 \times 0{\cdot}95 = 28{\cdot}5 \Rightarrow \sigma = \sqrt{28{\cdot}5} = 5{\cdot}3$$

We want x_1 such that

$$P(X < x_1) > 0{\cdot}90$$

$$\Rightarrow \quad P(Z < z_1) > 0{\cdot}90 \qquad \left[z_1 = \frac{x_1 - 30}{5{\cdot}3} \right]$$

$$\Rightarrow \quad P(0 < Z < z_1) > 0{\cdot}40$$

$$\Rightarrow \quad z_1 > 1{\cdot}28 \qquad \text{[From Normal Probability Tables]}$$

$$\Rightarrow \quad \frac{x_1 - 30}{5{\cdot}3} > 1{\cdot}28 \quad \Rightarrow \quad x_1 > 30 + 5{\cdot}3 \times 1{\cdot}28$$

$$\Rightarrow \quad x_1 > 30 + 6{\cdot}784 \quad \Rightarrow \quad x_1 > 36{\cdot}784 \simeq 37$$

Example 8·20 (*a*). *In a distribution exactly normal, 7% of the items are under 35 and 89% are under 63. What are the mean and standard deviation of the distribution ?*

(*b*) *Of a large group of men, 5% are under 60 inches in height and 40% are between 60 and 65 inches. Assuming a normal distribution, find the mean height and standard deviation.*

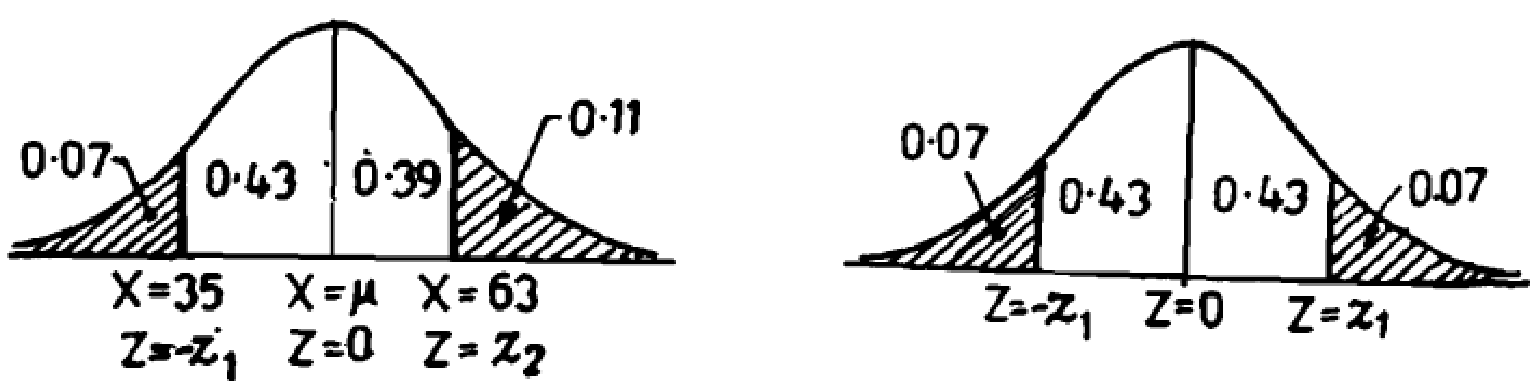
Solution. If $X \sim N(\mu, \sigma^2)$, then we are given

$$P(X < 63) = 0{\cdot}89 \quad \Rightarrow \quad P(X > 63) = 0{\cdot}11 \text{ and } P(X < 35) = 0{\cdot}07$$

The points $X = 63$ and $X = 35$ are located as shown in Fig. (*i*) below.

Since the value $X = 35$ is located to the left of the ordinate at $X = \mu$, the corresponding value of Z is negative.

When $X = 35$, $Z = \dfrac{35 - \mu}{\sigma} = -z_1$, (say),



and when $X = 63$, $Z = \dfrac{63 - \mu}{\sigma} = z_2$, (say).

Thus we have, as is obvious from figures (i) and (ii)

$$P(0 < Z < z_2) = 0 \cdot 39 \text{ and } P(0 < Z < z_1) = 0 \cdot 43$$

Hence from normal tables, we have

$$z_2 = 1 \cdot 23 \text{ and } z_1 = 1 \cdot 48$$

$$\therefore \qquad \frac{63 - \mu}{\sigma} = 1 \cdot 23 \text{ and } \frac{35 - \mu}{\sigma} = -1 \cdot 48$$

Subtracting, we get

$$\frac{28}{\sigma} = 2 \cdot 71 \Rightarrow \sigma = \frac{28}{2 \cdot 71} = 10 \cdot 33$$

$$\therefore \qquad \mu = 35 + 1 \cdot 48 \times 10 \cdot 33 = 35 + 15 \cdot 3 = 50 \cdot 3$$