Normal Curve and Sampling Distribution

<div align="right">

10

</div>

10.1 Normal Distribution

Normal distribution was first described by Abraham De Moivre and then developed by Laplace and Gauss. It is the most important of the theoretical distributions. It is a continuous probability distribution in which the random variable "X" can assume either a finite set of value or numerous infinite set of values. Distribution plays a vital role where inferences are made regarding the value of population mean (μ). A continuous variable "X" is said to be normally distributed if it has the probability density function represented by the equation:

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

where $\pi = 3.1416$ and $e = 2.7183$.

We know that $\frac{X-\mu}{\sigma} = Z$ (standard normal variate):

$$\text{So,} \, P(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(Z)^2}{2\sigma^2}}$$

10.2 Properties of Normal Curve

1. Normal distribution is a continuous probability distribution having two parameters: mean ($\bar{X}$) and standard deviation (σ).
2. Normal curve is a bell-shaped curve.
3. It is a symmetrical curve.
4. In this mean, median, and mode coincide, i.e., mean ($\bar{X}$) = median ($\tilde{X}$) = mode ($\widehat{X}$).
5. Median is at equal distance from Q_1 and Q_3, i.e., $\left(Q_3 - \tilde{X}\right) = \left(\tilde{X} - Q_1\right)$.
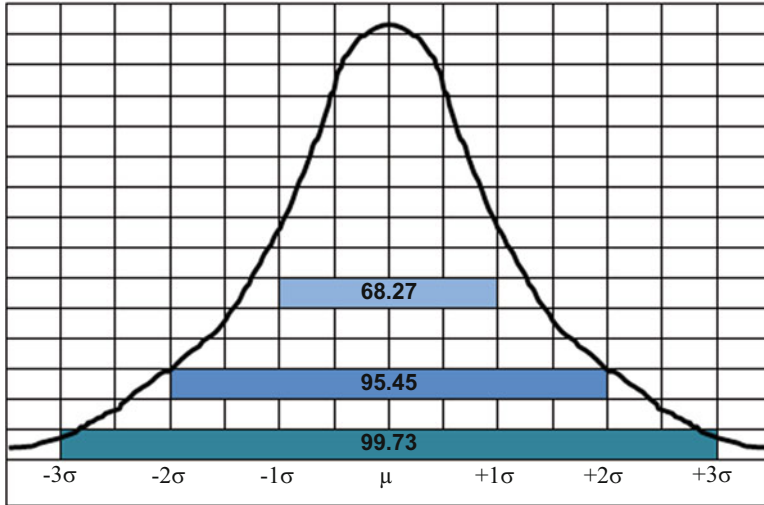
Fig. 10.1 Normal curve

6. It is unimodal class.
7. Its points of inflection are always at "one standard deviation" from the mean ($\bar{X}$).
8. The curve is symmetrical and does not touch the baseline.
9. The normal curve has maximum height at mean value.
10. The ordinate divides the curve into two equal parts.
11. The curve has "permanent areas relationships" as exhibited in Fig. 10.1.

10.3   Applications of Normal Curve

If X is normally distributed with mean μ and variance σ^2, the "standard normal variate" "Z" is normally distributed with mean "0" and variance 1. Area contained under the normal curve at any point "Z" could be noted down from the table for "Z values." Making use of this table, we can determine the area contained between any two ordinates in the normal curve.

10.4   Sampling Variations

Assume that an unbiased random sample is drawn from a population. The sample mean obtained may not exactly tally with the population mean. The error between the sample mean and the universe's mean is subjected to only chance fluctuations. This error is called "sampling error."

   If repeated samples of same size are drawn randomly from the same universe, not only the samples' means differ with the population mean, but also the sample means differ among each other. But there is a scope to prove that the distribution of these

samples' means follows a normal distribution with reference to population mean (μ), and "mean variance" σ^2/n (population variance divided by sample size), if the size of sample "n" is large ($n > 30$). If "n" is small, the sampling distribution of $\bar{X}$ follows the pattern of Student's "t"-distribution. This measure of variation ($2\sigma/\sqrt{n}$) differs from individual σ and is termed as "standard error (SE) of mean." SE is applied to calculate "t" value. For sampling distribution of means of large samples, the properties of normal distribution hold good.

Because:

$\mu \pm \sigma/\sqrt{n}$ contains 68% of the means.
$\mu \pm 2\sigma/\sqrt{n}$ contains 95% of the means.
$\mu \pm 3\sigma/\sqrt{n}$ contains 99.7% of the means.

10.5   Effect of Sample Size on Standard Error

It has been observed that with the increase in sample size, the reliability of the sample mean ($\bar{X}$) increases. In other words, the population mean (μ) can be estimated with greater confidence or lesser scope of error as the size of sample increases.

In a population study, standard deviation (σ) was 3.00 and mean ((μ) was 27.00. This gave out "variance" (σ^2) as 9.00. When six samples of sample size (n) of 4, 9, 16, 36, 100, and 400 were studied from the same population, the standard error ($\sigma/\sqrt{n}$) came out to be 1.50, 1.00, 0.75, 0.50, 0.30, and 0.15, respectively, thus reducing the sample observations' limits to be included in 68.27% distribution. The same has been illustrated in Table 10.1.

10.6   Assumptions

We always assume that:

1. Sample is random.
2. Sample has been drawn from the normal population.
3. If the population is normal, then the sampling distribution of means is also normal.
4. Even if the population is not much deviating from normality, the sampling distribution approaches normality with increasing size of sample.

If the population is markedly deviated from normality, then the sampling distribution will not follow the normal distribution. So, investigators should have knowledge of the structure of the population from which the sample is drawn. However, the assumption of normality could be often made without serious error in the absence of knowledge of the population.

Table 10.1 Effect of sample size on standard deviation on sampling distributions of means ($\mu = 27.00$ and $\sigma = 3.00$)

Sample size "n"	Standard error ($\sigma/\sqrt{n}$)	Limits to be included in 68.27 distribution: $\mu \pm (\sigma/\sqrt{n})$
4	1.50	25.50–28.50
9	1.00	26.00–28.00
16	0.75	26.25–27.75
36	0.50	26.50–27.50
100	0.30	26.70–27.30
400	0.15	26.85–27.15

In actual experience, the population is not exactly known. So, the mean and standard deviation of population cannot be known. Hence, the estimates obtained from a single sample are used to determine the mean and standard error of the sampling distribution of mean. The estimates from the sample, mean ($\bar{X}$) and standard deviation (s), are unbiased estimates of the population parameters μ and σ. Therefore, the "standard error of the mean" for a sample of size "n" is determined by $s/\sqrt{n}$.

Example 1 Suppose a random sample of size 100 is selected and excretion of urea in urine is detected in every individual. The mean urea excretion is found to be 8.000 g with a standard deviation of 1.600 g. When we workout standard error from these estimates, that would be $s/\sqrt{n} = \frac{1.6}{\sqrt{100}} = 0.16$. The confidence limits for the population mean can be determined as given below:

$$95\% \text{ Confidence limits} : \bar{X} \pm 2\,\text{SE} = 8.0 + 2 \times 0.16 = 8.0 \pm 0.32$$

This means the range 7.68–8.32 g would contain the population mean with 95% confidence. In other words, if 100 repeated samples are drawn of the same population, 95% individuals would have the value of excreted urea between 7.68 and 8.32 g.