Econometrics- Unit-1

Normal distribution

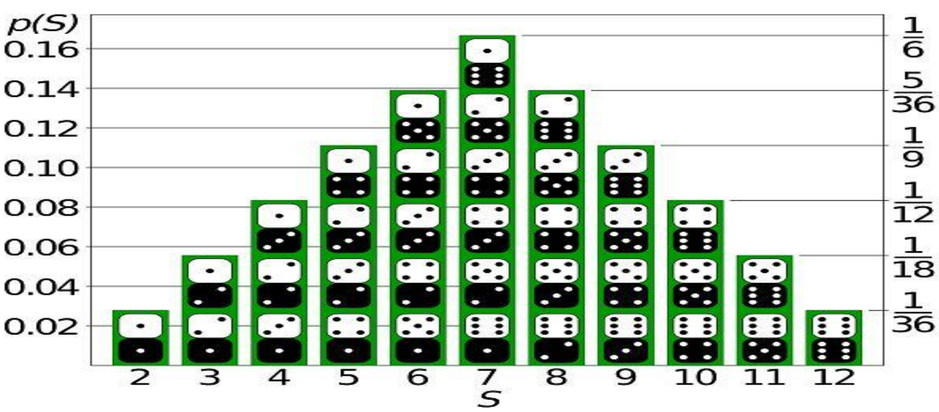
The normal distribution is widely used in understanding distributions of factors in the population. Normal/Gaussian Distribution is a bell-shaped graph which encompasses two basic terms- mean and standard deviation. Normal distribution follows the central limit theory which states that various independent factors influence a particular trait. When these all independent factors contribute to a phenomenon, their normalized sum tends to result in a Gaussian distribution. The mean of the distribution determines the location of the center of the graph, and the standard deviation determines the height and width of the graph and the total area under the normal curve is equal to 1. The bell curve is symmetrical. Half of the data will fall to the left of the mean; half will fall to the right. A **normal** (or **Gaussian** or **Gauss** or **Laplace–Gauss**) **distribution** is a type of continuous probability distribution for a real-valued random variable. The general form of its probability density function* is

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

The parameter is the mean or expectation of the distribution (and also its median and mode),while the parameter is its standard deviation. The variance of the distribution is. A random variable with a Gaussian distribution is said to be normally distributed, and is called a normal deviate.

Example of Normal Distribution:-

A fair rolling of dice is also a good example of normal distribution. In an experiment, it has been found that when a dice is rolled 100 times, chances to get '1' are 15-18% and if we roll the dice 1000 times, the chances to get '1' is, again, the same, which averages to 16.7% (1/6). If we roll two dices simultaneously, there are 36 possible combinations. The probability of rolling '1' (with six possible combinations) again averages to around 16.7%, i.e., (6/36). More the number of dices more elaborate will be the normal distribution graph.



Uses of Normal Distribution:

Normal distributions are important in statistics and are often used in the natural and social sciences to represent real-valued random variables whose distributions are not known. Their importance is partly due to the central limit theorem. It states that, under some conditions, the average of many samples (observations) of a random variable with finite mean and variance is itself a random variable—whose distribution converges to a normal distribution as the number of samples increases. Therefore, physical quantities that are expected to be the sum of many independent processes, such as measurement errors, often have distributions that are nearly normal.

Parameters of Normal Distribution

The two main parameters of a (normal) distribution are the mean and standard deviation. The parameters determine the shape and probabilities of the distribution. The shape of the distribution changes as the parameter values change.

1. Mean

The mean is used by researchers as a measure of central tendency. It can be used to describe the distribution of variables measured as ratios or intervals. In a normal distribution graph, the mean defines the location of the peak, and most of the data points are clustered around the mean. Any changes made to the value of the mean move the curve either to the left or right along the X-axis.

2. Standard Deviation

The standard deviation measures the dispersion of the data points relative to the mean. It determines how far away from the mean the data points are positioned and represents the distance between the mean and the observations. On the graph, the standard deviation determines the width of the curve, and it tightens or expands the width of the distribution along the x-axis. Typically, a small standard deviation relative to the mean produces a steep curve, while a large standard deviation relative to the mean produces a flatter curve.

Properties

All forms of (normal) distribution share the following characteristics:

1. It is symmetric

A normal distribution comes with a perfectly symmetrical shape. This means that the distribution curve can be divided in the middle to produce two equal halves. The symmetric shape occurs when one-half of the observations fall on each side of the curve.

2. The mean, median, and mode are equal

The middle point of a normal distribution is the point with the maximum frequency, which means that it possesses the most observations of the variable. The midpoint is also the point where these three measures fall. The measures are usually equal in a perfectly (normal) distribution.

3. Empirical rule

In normally distributed data, there is a constant proportion of distance lying under the curve between the mean and specific number of standard deviations from the mean. For example, 68.25% of all cases fall within +/- one standard deviation from the mean. 95% of all cases fall within +/- two standard deviations from the mean, while 99% of all cases fall within +/- three standard deviations from the mean.

4. Skewness and kurtosis

Skewness and kurtosis are coefficients that measure how different a distribution is from a normal distribution. Skewness measures the symmetry of a normal distribution while kurtosis measures the thickness of the tail ends relative to the tails of a normal distribution.

*Probability Density Function:- In probability theory, a probability density function (PDF), or density of a continuous random variable, is a function whose value at any given sample (or point) in the sample space (the set of possible values taken by the random variable) can be interpreted as providing a *relative likelihood* that the value of the random variable would equal that sample.

Example- Suppose bacteria of a certain species typically live 4 to 6 hours. The probability that a bacterium lives *exactly* 5 hours is equal to zero. A lot of bacteria live for approximately 5 hours, but there is no chance that any given bacterium dies at exactly 5.0000000000... hours. However, the probability that the bacterium dies between 5 hours and 5.01 hours is quantifiable. Suppose the answer is 0.02 (i.e., 2%). Then, the probability that the bacterium dies between 5 hours and 5.001 hours should be about 0.002, since this time interval is one-tenth as long as the previous. The probability that the bacterium dies between 5 hours and 5.0001 hours should be about 0.0002, and so on.

In these three examples, the ratio (probability of dying during an interval) / (duration of the interval) is approximately constant, and equal to 2 per hour (or 2 hour^{-1}). For example, there is 0.02 probability of dying in the 0.01-hour interval between 5 and 5.01 hours, and (0.02 probability / 0.01 hours) = 2 hour^{-1}. This quantity 2 hour^{-1} is called the probability density for dying at around 5 hours. Therefore, the probability that the bacterium dies at 5 hours can be written as (2 hour^{-1}) dt. This is the probability that the bacterium dies within an infinitesimal window of time around 5 hours, where dt is the duration of this window. For example, the probability that it lives longer than 5 hours, but shorter than (5 hours + 1 nanosecond), is (2 hour^{-1})×(1 nanosecond) ≈ 6×10^{-13} (using the unit conversion 3.6×10^{12} nanoseconds = 1 hour).

There is a probability density function *f* with *f*(5 hours) = 2 hour^{-1}. The integral of *f* over any window of time (not only infinitesimal windows but also large windows) is the probability that the bacterium dies in that window.