Normal Distribution

Mahuya Maity[1], Papiya Saha[2]

[1]Faculty, College of Nursing, NRSMc & H, Kolkata, India

[2]Faculty, College of Nursing, NRSMc & H, Kolkata, India

Abstract: ***The primary purpose of this paper is to discuss the idea of normal distribution. The authors describe the definition, formula properties and applications of normal distribution. Normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graphical form, the normal distribution appears as a "bell curve". That means, the standard normal distribution has two parameters: The mean and the standard deviation (1).This article focuses on properties, formula and its application in nursing research. The normal distribution is one type of symmetrical distribution. Symmetrical distributions occur when where a dividing line produces two mirror images. Not all symmetrical distributions are normal since some data could appear as two humps or a series of hills in addition to the bell curve that indicates a normal distribution. For all normal distributions, 68.2% of the observations will appear within plus or minus one standard deviation of the mean; 95.4% of the observations will fall within +/- two standard deviations; and 99.7% within +/- three standard deviations. This fact is sometimes referred to as the "empirical rule," a heuristic that describes where most of the data in a normal distribution will appear. This means that data falling outside of three standard deviations ("3- sigma") would signify rare occurrences.***

Keywords: Symmetrical distribution, mean, standard deviation, bell curve

1. Properties of the Normal Distribution

The normal distribution has several key features and properties that define it.

First, its mean (average), median (midpoint), and mode (most frequent observation) are all equal to one another. Moreover, these values all represent the peak, or highest point, of the distribution. The distribution then falls symmetrically around the mean, the width of which is defined by the standard deviation.

The representation of data is inclusive of two parameters: *The measure of central tendency and the measure if dispersion*. The measure of central tendency is direction towards the central most value of the data set as given by the mean or median. The measure of dispersion includes standard deviation (SD), standard error and confidence interval.[2]

The distribution of data is again dependent on the data type. In case of categorical data the distribution is binomial as the out come is binary. *E.g. Present/Absent; Yes/No; Normal/Diseased*. However, with continuous data, there is distribution of data on either side of the mean (measure of central tendency) as given by SD (measure of dispersion). When this distribution follows a bell-shape, then it is called *normal*.[2]

Skewness

Skewness measures the degree of symmetry of a distribution. The normal distribution is symmetric and has a skewness of zero.

If the distribution of a data set instead has a skewness less than zero, or negative skewness (left-skewness), then the left tail of the distribution is longer than the right tail; positive skewness (right-skewness) implies that the right tail of the distribution is longer than the left.

Kurtosis

Kurtosis measures the thickness of the tail ends of a distribution in relation to the tails of a distribution. The normal distribution has a kurtosis equal to 3.0.

Distributions with larger kurtosis greater than 3.0 exhibit tail data exceeding the tails of the normal distribution (e.g., five or more standard deviations from the mean). This excess kurtosis is known in statistics as leptokurtic, but is more colloquially known as "fat tails." The occurrence of fat tails in financial markets describes what is known as tail risk.

Distributions with low kurtosis less than 3.0 (platykurtic) exhibit tails that are generally less extreme ("skinnier") than the tails of the normal distribution.

The Formula for the Normal Distribution

The normal distribution follows the following formula. Note that only the values of the mean (μ) and standard deviation (σ) are necessary

Normal Distribution Formula.

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

where:

- x = value of the variable or data being examined and f(x) the probability function
- μ = the mean
- σ = the standard deviation

2. Applications

In most cases, the distribution of data from biomedical research is asymmetric. It should also be noted that a normal distribution can take place only for continuous quantitative variables. If the trait in question is a qualitative, ordinal, or even quantitative discrete, its distribution cannot be normal. For instance, the number of family members, the number of antibiotics prescribed, the number of changes in antibacterial therapy, the number of rooms in the room, and other examples. Statistical processing of such characteristics should be performed by using nonparametric methods.

When taking such variables as height and weight, the latter will probably have a larger standard deviation. Growth is the value that, as a rule, does not depend on a person, and the weight of many people is dynamic. In case of any shift in the mode of life, the body mass parameter may vary by the influence of external circumstances, and its values may vary to a greater or lesser extent. Therefore, such a variable is considered dynamic with a larger standard deviation.[3]

The normal distributions are closely associated with many things such as:
- Marks scored on the test
- Heights of different persons
- Size of objects produced by the machine
- Blood pressure and so on.

Statistical notes:
- The parameters of normal distribution are mean and SD.
- Distribution is a function of SD.
- Sample size plays a role in normal distribution.
- Skewed distribution can also be representative if the population under study.
- Normal distribution of data can be ascertained by certain statistical tests.

3. Conclusion

The normal distribution, also known as the Gaussian distribution, is the most important probability distribution in statistics for independent, random variables. Most people recognize its familiar bell-shaped curve in statistical reports. It is the most important probability distribution in statistics because it accurately describes the distribution of values for many natural phenomena. Characteristics that are the sum of many independent processes frequently follow normal distributions. For example, heights, blood pressure, measurement error, and IQ scores follow the normal distribution.

References

[1] Sheldom M. Ross. "Introductory Statistics," Section 7.4. Academic Press, 2017.
[2] Krithikadatta J, Valarmathi S. Research Methodology in dentistry: Part II — The relevance of statistics in research. *J Conserv Dent.* 2012;15:206–213. [PMC free article] [PubMed] [Google Scholar
[3] Jamali, R., Arj, A., Razavizade, M., & Aarabi, M. H. (2016). Prediction of nonalcoholic fatty liver disease via a novel panel of serum adipokines. *Medicine*, *95*(5), e2630. Web.