SAGE reference

The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation

Normal Distribution

Contributors: Dorothy J. Musselwhite & Brian C. Wesolowski
Edited by: Bruce B. Frey
Book Title: The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation
Chapter Title: "Normal Distribution"
Pub. Date: 2018
Access Date: February 26, 2018
Publishing Company: SAGE Publications, Inc.
City: Thousand Oaks,
Print ISBN: 9781506326153
Online ISBN: 9781506326139
DOI: http://dx.doi.org/10.4135/9781506326139.n476
Print pages: 1155-1157

The normal distribution is a hypothetical symmetrical distribution used to make comparisons among scores or to make other kinds of statistical decisions. The shape of this distribution is often referred to as "bell shaped" or colloquially called the "bell curve." This shape implies that the majority of scores lie close to the center of the distribution, and as scores drift from the center, their frequency decreases.

Normal distributions belong to the family of continuous probability distributions or probability density functions. A probability density function is a function meant to communicate the likelihood of a random variable to assume a given value. This function is graphed by plotting the variable, *x*, by the probability of that variable occurring, *y*. These normal probability distributions are characterized by the aforementioned symmetric bell shape but can have any real mean, labeled μ, and any positive real standard deviation, labeled σ. Specifically, the normal distribution is characterized by continuous data, meaning the data can occupy any range of values. In special cases, the normal distribution can be standardized in which the mean becomes 0 and the standard deviation becomes 1. All normal distributions can be transformed or standardized to the standard normal distribution.

The normal distribution is commonly named the "Gaussian distribution," after Carl Friedrich Gauss, a German mathematician who made significant advancements of statistical concepts. Less frequently, the normal distribution may be called the "Laplace distribution," after Pierre-Simon Laplace. The remainder of this entry reviews the history of normal distribution, explains the defining function of normal distribution, explores its properties, highlights the differences between normal distribution and standard normal distribution, and reviews assumptions and tests of normality.

History

The first affiliation with normal distribution stemmed from errors of measurement. Galileo Galilei looked specifically within astronomy to notice that the errors in observations were not random. Small errors far outweighed the larger errors, and these errors had a tendency to be symmetrically distributed around a peak value.

In 1895, Karl Pearson is credited with the first appearance of the term *normal distribution* from his seminal paper. However, the term also appeared in work by Charles Peirce in 1783, Francis Galton in 1889, and Henri Poincaré in 1893. The first mathematical derivation of the normal distribution is attributed to Abraham DeMoivre in his *Approximatio ad summam terminorum binomii* $(a + b)^n$ *in seriem expansi*. DeMoivre used integral calculus to estimate a continuous distribution, resulting in a bell-shaped distribution.

In 1808, Robert Adrain, an American mathematician, debated the validity of the normal distribution, expounding on distributions of measurement errors. His discoveries led to further work in proving Adrien-Marie Legendre's method of least squares. In 1809, without knowledge of Adrain's work, Gauss published his *Theory of Celestial Movement*. This work presented substantial contributions to the statistics field, including the method of least squares, the maximum likelihood parameter estimation, and the normal distribution. The significance of these contributions is possibly why Gauss is given credit over Adrain in regard to the normal distribution. In use from 1991 to 2001, the German 10 DM banknote displayed a portrait of Gauss and a graphical display of the normal density function.

In the early 1800s, Adolphe Quetelet, Walter Weldon, and Pearson worked to apply the concept of the normal distribution to the biological and social sciences, eventually cofounding the journal *Biometrika*. In 1994, American psychologist Richard Herrnstein and political scientist Charles Murray published *The Bell Curve: Intelligence and Class Structure in American Life*. This publication led to the term *bell curve* becoming a more widely known concept. Herrnstein and Murray looked at the relation between intelligence scores and social outcomes, resulting in implications of an ever-increasing social stratification based on intelligence.

Definition

The normal distribution is constructed using the normal density function:

$$px = e - x - \mu^2/2\sigma^2\sigma^2\pi.$$

This exponential function is comprised of a constant (*e*), the mean (μ), the standard deviation (σ), and the variance (σ^2). The formula is often shortened to $N(\mu, \sigma^2)$. If $N(0, 1)$, so that $\mu = 0$ and $\sigma^2 = 1$, the resulting distribution is the standard normal distribution.

The shape of the normal distribution is based on two parameters: the mean (μ) and the standard deviation (σ). The mean controls the *x*-axis, and the standard deviation controls the *y*-axis. The mean influences the location of the apex of the distribution and is therefore called the location parameter. The variance (σ^2) influences how wide the distribution appears and is therefore called the scale parameter. A larger variance will result in a wider bell curve.

Properties

In a normal distribution, the curve is entirely symmetrical around the mean, such that $x = \mu$. This symmetrical distribution shows that with the mean, the median and mode must also coincide. There are more data observations closer to these central tendency values than to the extremes of the bell shape. Along the *y*-axis, the graph stretches from −∞ to +∞. The normal distribution, however, is a purely hypothetical model. Rarely do observations exist in the world that fit the model perfectly. Rather, scores and distributions come close to the normal distribution. As the sample size increases, the distribution becomes closer to the hypothetical model.

The function approximates the number of observations that should fall within specific areas of the curve. For example, within 1 standard deviation of the mean (+1 and −1), approximately 68.3% of observations should appear here. Looking specifically at 1 standard deviation above the mean, only 34.1% of observations fall between the mean and the value of the mean plus 1 standard deviation above that mean. Roughly, 95.4% of observations should fall within 2 standard deviations from the mean (+2 and −2). Finally, about 99.7% of observations fall within 3 standard deviations from the mean (+3 and −3).

The normal curve can be used to determine the percentage of scores above or below a certain score. Specifically, between the mean and 3 standard deviations above the mean, approximately 50% of observations should occur in this interval (34.13 + 13.59 + 2.14).

Skewness and kurtosis are two ways a distribution can deviate from the normal, idealized shape. In a normal distribution, skew and kurtosis have values of 0. As the distribution deviates further from normal, these values move above and below 0. Skewness refers to a lack of symmetry, where most of the scores are gathered at one end of the scale. Positively skewed refers to a distribution with a cluster of scores at the lower end of the scale with the tail at the higher, more positive end of the scale. Negatively skewed refers to a distribution with the tail at the lower end of the scale, and the cluster of scores positioned at the higher, positive end.

Kurtosis refers to the pointiness of the distribution. More specifically, kurtosis describes the degree to which scores cluster at the ends of the distribution or the tails. Positive kurtosis is a pointier version of the normal distribution, with many scores in the tails. It is often called a heavy-tailed distribution or leptokurtic. Negative kurtosis, often called platykurtic, is flatter than the normal distribution and is relatively thin in the tails.

Standard Normal Distribution

The standard normal distribution is a version of the normal distribution in which the normal random variable has a mean of 0 and a standard deviation of 1. In the standard distributions, the random variables are transformed into *z* scores using the following formula for use with a population:

$$z = (X - \mu)\,\sigma,$$

where *X* is the normal random variable, μ is the mean of the data, and σ is the standard deviation. The following formula for *z* scores is used with a sample:

$$z = (X - X)\,s,$$

where *X* is the normal random variable (or score), *X* is the mean of the data, and *s* is the standard deviation. Most frequently in use with a standard distribution is the standard normal distribution table, which dictates cumulative probability based on the *z* score calculated. This table gives values of the area under each part of the curve at the value of *z*. The areas are related to probability. Only in a standardized normal distribution does the total area under the curve equal one (1.0). The area above the *z* score indicates the likelihood of those values occurring, and the area below the *z* score indicates the likelihood of those values occurring.

Assumptions of Normality

In the interpretation of data, it is important that all evidence is evaluated objectively or free of bias. Outliers can directly affect this interpretation as well as any violation of the assumption of normality. This assumption is one of a varied list of assumptions of statistical tests but relates directly to the normal distribution.

Extreme scores may bias estimates of parameters. Specifically, the mean may be influenced more by outliers than the median. Confidence intervals are based on parameter estimates and therefore are also influenced by the bias of outliers. To achieve accurate confidence intervals, estimates must come from a normal distribution. Null hypothesis significance testing assumes that parameter estimates are normally distributed because other test statistics, such as those from the *t* test, *F* test, and χ^2 distribution, are normally distributed. Because populations are often unavailable for testing, significance tests will be accurate when the sampling distribution is normally distributed.

If the sample size is large enough, however, the assumption of normality becomes less of a concern. A larger sample size increases the normality of the distribution and therefore will result in more accurate confidence intervals, significance tests, and estimates of parameters. The definition of “large” will vary from distribution to distribution. The most generally accepted value for sample size is 30, but skew and kurtosis can also impact how large this value should be. Sample sizes upward of 100 may be necessary to achieve a more accurate sampling distribution.

The misunderstanding that occurs most frequently with the assumption of normality is that the data alone need to be normally distributed, which is not the case. The errors, or residuals, of the data should be normally distributed as well as the sampling distribution. However, the raw data are likely to have a varying shape.

Tests of Normality

Many parametric tests are based on the assumption of normality, which assumes the sampling distribution of the population parameter is normally distributed. This assumption does not mean that the sample data being analyzed should be normally distributed.

Two tests of normality exist to compare scores in a sample to a normally distributed set of scores with identical mean and standard deviation. A significant p value ($p < .05$) from these tests indicates that the distribution is significantly different from a normal distribution. The Kolmogorov–Smirnov test and the Shapiro–Wilk test both test for this significance. However, the Shapiro–Wilk test has more power to detect differences from normality. Therefore, the Shapiro–Wilk test may have significant values when the Kolmogorov–Smirnov test does not. These tests should be used carefully, as false significance may occur when testing larger samples. Both tests should be used simultaneously with histograms or plots and the aforementioned values of skew and kurtosis.

See alsoKurtosis; Skewness; Standard Deviation; Standard Error of Measurement; Variance

Dorothy J. MusselwhiteBrian C. Wesolowski
http://dx.doi.org/10.4135/9781506326139.n476
10.4135/9781506326139.n476

Further Readings

Boyle, J. D., & Radocy, R. E. (1987). Measurement and evaluation of musical experiences. New York, NY: Schirmer Books.
Field, A. (2013). Discovering statistics using IBM SPSS statistics. London, UK: SAGE.
Gravetter, F. J., & Wallnau, L. B. (2011). Essentials of statistics for the behavioral sciences. Belmont, CA: Wadsworth.
Gross, J. (2004). A normal distribution course. Frankfurt, Germany: Peter Lang.
Kubiszyn, T., & Borich, G. (2003). Educational testing and measurement: Classroom application and practice. New York, NY: Wiley.
Payne, D. A. (2003). Applied educational assessment. Toronto, Canada: Wadsworth.
Reid, H. M. (2014). Introduction to statistics: Fundamental concepts and procedures of data analysis. Thousand Oaks, CA: SAGE.