Print

Normal (Gaussian) Distribution

Because sampling involves variability, parameter estimates also have variability. Often, the variability of sample statistics is approximately normal.

Perhaps the most famous statistical theorem, the central limit theorem, indicates that the sample mean of ANY distribution approximates a normal distribution. That is, if we could sample all subsets of our population of size n, and create a histogram, the sample means will be bell shaped with the population mean and standard error measuring its variability.

For example, imagine our population distribution is bell shaped, with a population mean, μ=50, and standard deviation, σ=5. Suppose we sample n=30 observations and the sample mean is 48.3 Then we take another 30 observation sample, and the sample mean is 52.1. If we repeat this process many times, the distribution of the sample mean will follow a bell shaped normal distribution with population mean μ=50 and standard error of σ divided by the square root of n=30.

Another name for the normal distribution is the Gaussian distribution. The Gaussian distribution is bell-shaped, symmetric, and defined by two parameters, μ (the mean) and σ (the standard deviation). The mean locates the midpoint of the distribution, or the peak of the bell, and the standard deviation describes the variability, or spread of the distribution. The area under the curve between two points is the probability of getting values between those two points.

Some well-known probabilities are associated with the mean and standard deviation of the distribution. For example, approximately 68% of the normal distribution lies within 1 standard deviation of the mean. Approximately 95% lies within 2 standard deviations of the mean, and approximately 99% lies within 3 standard deviations.

Statisticians often consider values that are more than 2 standard deviations from the mean as unusual. And now you can see why. Only about 5% of all values are that far away from the mean. Depending on the context, some statisticians treat only values more than 3 standard deviations away from the mean as unusual.

Because the normal distribution has many useful mathematical properties, statistical procedures for data based on a random sample often assume the normal distribution. So it's important to know how to check this assumption for your data.

Close