Nonparametric Statistics

What Are Nonparametric Statistics?

Nonparametric statistics refer to a statistical method in which the data is not required to fit a normal distribution. Nonparametric statistics uses data that is often ordinal, meaning it does not rely on numbers, but rather on a ranking or order of sorts. For example, a survey conveying consumer preferences ranging from like to dislike would be considered ordinal data.

Nonparametric statistics includes nonparametric descriptive statistics, statistical models, inference, and statistical tests. The model structure of nonparametric models is not specified *a priori* but is instead determined from data. The term *nonparametric* is not meant to imply that such models completely lack parameters, but rather that the number and nature of the parameters are flexible and not fixed in advance. A histogram is an example of a nonparametric estimate of a probability distribution.

Understanding Nonparametric Statistics

In statistics, parametric statistics includes parameters such as the mean, median, standard deviation, variance, etc. This form of statistics uses the observed data to estimate the parameters of the distribution. Under parametric statistics, data is assumed to fit a normal distribution with unknown parameters μ (population mean) and σ^2 (population variance), which are then estimated using the sample mean and sample variance.

Nonparametric statistics does not assume that data is drawn from a normal distribution. Instead, the shape of the distribution is estimated under this form of statistical measurement. While there are many situations in which a normal distribution can be assumed, there are also some scenarios in which it will not be possible to determine whether the data will be normally distributed.

Examples of Nonparametric Statistics

In the first example, consider a researcher that wants an estimate of the number of babies in North America born with brown eyes may decide to take a sample of 150,000 babies and run an analysis on the data set. The measurement that they derive will be used as an estimate of the entire population of babies with brown eyes born the following year.

For a second example, consider a different researcher who wants to know whether going to bed early or late is linked to how frequently one falls ill. Assuming the sample is chosen randomly from the population, the sample size distribution of illness frequency can be assumed to be normal. However, an experiment that measures the resistance of the human body to a strain of bacteria cannot be assumed to have a normal distribution.

This is because a randomly selected sample data may be resistance to the strain. On the other hand, if the researcher considers factors such as genetic makeup and ethnicity, he may find that a

sample size selected using these characteristics may not be resistant to the strain. Hence, one cannot assume a normal distribution.

This method is useful when the data has no clear numerical interpretation and is best to use with data that has a ranking of sorts. For example, a personality assessment test may have a ranking of its metrics set as strongly disagree, disagree, indifferent, agree, and strongly agree. In this case, nonparametric methods should be used.

Special Considerations

Nonparametric statistics have gained appreciation due to their ease of use. As the need for parameters is relieved, the data becomes more applicable to a larger variety of tests. This type of statistics can be used without the mean, sample size, standard deviation, or the estimation of any other related parameters when none of that information is available.

Since nonparametric statistics makes fewer assumptions about the sample data, its application is wider in scope than parametric statistics. In cases where parametric testing is more appropriate, nonparametric methods will be less efficient. This is because the results obtained from nonparametric statistics have a lower degree of confidence than if the results were obtained using parametric statistics.