What is nonparametric statistics?

We know that the basis of inferential statistics is parameter estimation, i.e. estimating the parameter of a population from information gained from a random sample presumed to have been drawn from that population. In most statistics discussion, the underlying distributions from which our samples selected are being of a particular type, such as the normal distribution, for inferences made from the relevant statistical tests to be valid; hence these techniques are called *parametric statistics*.

What about the scenarios in which we know or suspect that the population does not meet the assumptions for a particular statistical test? Also, we have been using tests with two or more samples with an assumption that the populations have equal variances. However, we know well that there can be a situation in which some of our population distributions are clearly non-normal and where we do not have equal variability.

In these cases, a different set of statistical techniques known as *nonparametric statistics* can be used. They are also sometimes called as *distribution-free statistics* because they make few or no assumptions about the underlying distribution of the data.

Nonparametric statistics are often applied to data sets in which data have been collected as ranks rather than as raw score values, or rank data are substituted for raw scores due to concerns about the distribution of the raw data. These rank data by definition is ordinal, which refers to data that has some meaningful order, so that higher values represent more of some characteristics than lower values, or objects are categorized from best to worst. However, they do not permit comparisons quantitatively in terms of how much larger, lower or better.

In here we state the hypotheses in a different form than for parametric tests and they do not have the normality assumption. Rather than testing that population means are different, as is the case with two sample tests or with ANOVA, we test that if the populations have the same distribution or if the position or rank of the population differs.

In nonparametric test procedures, the data is often replaced by signs or ranks or both signs and ranks. The actual data itself is not analyzed, however. Hence, there are comments that nonparametric tests are wasteful of information because the original data is not used directly in reaching statistical inferences about populations, but rather signs and ranks are used in reaching these conclusions.

There are many nonparametric test statistics, including the popular chi-square test, McNemar's test, the Wilcoxon rank sum test and the Wilcoxon signed ranked test for pair difference experiment. We shall discuss some of these tests in the future blogs.

Generally speaking, nonparametric techniques are more robust than their parametric counterparts, because they are less influenced by departures from model assumptions or unusual values (such as outliers), but by nature are less powerful than their parametric equivalents. For this reason, if our data meet the assumptions for a parametric test, we should carry out the parametric test in the first instance; if not, we use the nonparametric equivalent.