NON-PARAMETRIC STASTISTICS

The term non-parametric was first used by Wolfowitz, 1942. To understand the idea of nonparametric statistics it is required to have a basic understanding of parametric statistics which we have already discussed. A parametric test requires a sample to be normally distributed. A nonparametric test does not rely on parametric assumptions like normality.

Nonparametric test create flexible demands of the data. To make standard parametric legitimate, some provisions need to fulfilled, especially for minor sample sizes. For example, the requirement of the one sample t-test is that the observation must be made from ordinarily distributed population. In case, the provision is defined, then the resultants may not be credible. However, in case of Wilcoxon Signed rank test to illustrate valid inference, normality is not required.

We can assume that the sampling distribution is normal even if we are not sure that the distribution of the variable in the population is normal, as long as our sample is large enough, (for example, 100 or more observations). However, if our selected sample is too large, then those teats can only be utilized if we are assured that the variable is disseminated normally.

The applications of tests that are based on the normality assumptions are restricted by the deficiency of accurate measurement. For example, a study measures Grade Point Average (GPA) in place of percentage Marks. This measurement scale does not measure the exact distance between the marks of two students. GPA allows us only to rank the students from "good" to "poor" students. This measurement is called the ordinal scale. Statistical techniques such as Analysis of Variance, t-test etc. assume that the data are measured either on interval or ratio scale. In such situations where data is measured on nominal or ordinal scale nonparametric tests are more useful.

Thus, nonparametric tests are used when either:

- Sample is not normally distributed.
- Sample size is small.
- The variables are measured on nominal or ordinal scale.

There is at least one nonparametric equivalent for each parametric general type of test. Broadly, these tests fall into the following categories:

- Test of differences between groups (independent samples)
- Test of differences (dependent samples)
- Test of relationships between variables.

The concepts and procedure to undertake Run Test, Chi-Square Test, Wilcoxon Signed Rank Test, Mann-Whitney Test and Kruskal-Wallis Test are discussed here under:

Run Test: Run Test is used to examine the randomness of data. Many statistical procedures require data to be randomly selected. Run Test can be explained with the example of tossing of a fair coin, where the probability of getting a head or tail is equal which is 0.5. Suppose we denote Head by "1" and tail by "0" and record the outcomes as shown below, we conduct Run Test to see whether the sample is randomly chosen or not. The null and alternative hypotheses are:

H_0: The sample is randomly selected.

H_1: The sample is not randomly selected.

The Run length is calculated by computing the number of 0's or 1's in sequence. For example the run length for 0 and 1 in the following sequence is 5 and 4 respectively and number of runs is 2.

000001111

We shall conduct Run test on our data set to examine whether the students belonging to different faculties (graduation) are randomly selected or not.

Chi-Square Test: Chi-Square Test is used to examine the association between two or more variables measured on categorical scales. Chi-Square is used most frequently to test the statistical significance of result reported in bivariate tables, and interpreting bivariate tables is integral to interpreting the results of a chi-square test.

Bivariate tabular (Cross tabulation) analysis is used when trying to summarize the intersections of independent and dependent variables and to examine the relationship (if any) between those variables. For example, to know if there is any association between the **Gender** and their **Location,** Chi-Square test can be applied. In this case our dependent variable is **Location**. We control the independent variable **Gender** and elicit as well as measure the dependent variable Location to test the hypothesis, whether there is some association between these two variables.

The Chi-Square test is a statistical technique to examine the association or statistical independence between the row and column variables in a two – way table. The null and alternative hypotheses for Chi-Square Test are:

H_0: There is no association between the row (Gender) and column (Location) variables.

H_1: There is association between the row (Gender) and column (Location) variables.

Many researchers often get confused between statistical significance and strength of the relationship. People tend to think that more significant (the lower the P-value) relationship means stronger relationship. The significance level is influenced by the strength of the relationship and sample size. We require different measure to capture the strength of the relationship, or the effect size.

Mann- Whitney U Test: Generally, the t-test for independent samples is used, if two samples are compared over their mean value for some variable interest. Nonparametric alternatives for the test are the Wald-Wolfowitz Run test, the Mann Whitney U test, and the Kolmogorov-Smirnov two sample test.

Mann Whitney U test compares the sums of ranks of two independent groups.

Wilcoxon Signed Rank Test: Wilcoxon Signed Rank Test (also known as Wilcoxon Matched Pair Test) is the non-parametric version of dependent sample t-test or paired sample t-test. Sign test is the other nonparametric alternative to the paired sample t-test. If the variables of interest are dichotomous in nature (Male and Female or Yes and No) then McNemar's Chi-Square test is used.

Wilcoxon Signed Rank Test is also a nonparametric version for one sample t-test. Wilcoxon Signed Rank Test compares the medians of the groups under two situations (paired samples) or it compares the median of the group with hypothesized median (one sample).

Kruskal –Wallis Test: Kruskal –Wallis Test is used with multiple groups. It is the non-parametric version of one-Way ANOVA. Median test is another nonparametric alternative to one-Way ANOVA. Kruskal –Wallis Test compares medians of more than two independent groups.