Non-parametric Test

A Non-parametric Test is a test that does not depend on the particular form of the basic frequency function from which the samples are drawn. In other words, non-parametric test does not make any assumptions regarding the form of the population.

However, certain assumptions associated with non-parametric tests are:
i. Sample observations are independent.
ii. The variable under study is continuous.
iii. P.d.f. is continuous.
iv. Lower order moments exist.

Advantages of non-parametric methods over parametric methods:
i. Non-parametric methods are readily comprehensible, very simple and easy to apply and do not require complicated sample theory.
ii. No assumption is made about the form of the frequency function of the parent population from which sampling is done.
iii. Since the socio-economic data are not, in general, normally distributed, non-parametric test have found applications in Psychometry, Sociology and Educational Statistics.
iv. Non-parametric tests are available to deal with the data which are given in ranks.

Disadvantages of non-parametric methods over parametric methods:
i. Non-parametric tests can be used only if the measurements are nominal or ordinal. Even in that case, if a parametric test exists it is more powerful than the non-parametric test.
ii. Non-parametric tests are designed to test statistical hypothesis only and not for estimating the parameters.

Chi-square tests: There are two main types of chi-square test. The chi-square test for goodness of fit applies to the analysis of a single categorical variable, and the chi-square test for independence or relatedness applies to the analysis of the relationship between two categorical variables.

Assumption testing: There are three assumptions you need to address before conducting chi-square tests.
1. **Random sampling-** observations should be randomly sampled from the population of all possible observations.
2. **Independence of observations-** each observation should be generated by a different subject and no subject is counted twice.
3. **Size of expected frequency-** when the number of cell is less than ten and particularly when the total sample size is small, the lowest expected frequency required for a chi-square test is five. However, the observed frequencies can be any value, including zero.

Chi-square test for goodness of fit: The test of significance involves the comparison between a set of actual and a set of theoretical frequencies, the latter having been obtained in accordance with the hypothesis we wish to test. This sort of test is known as a test of goodness of fit. The calculation of theoretical frequencies, given the form of the hypothetical distribution, on the basis of observed frequency table is often referred to as the fitting of a distribution.

Let the observed frequencies be denoted by $O_i \ (i = 1, 2, \ldots\ldots, k)$ and the expected frequencies, as determined by the null hypothesis, by $E_i \ (i = 1, 2, \ldots\ldots, k)$. Then the test statistic used is

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = \sum_{i=1}^{k} \frac{O_i^2}{E_i} - N$$

Where $N = \sum_{i=1}^{k} O_i = \sum_{i=}^{k} E_i$

This test statistic is distributed approximately as χ^2 with $k-r-1$ d.f., where *r* denotes the number of independent parameters estimated from the sample data in order to evaluate the expected frequencies.

Example: The following table outlines the attitudes of 60 people towards US military bases in Australia.

Attitude towards US military	**Frequency**
In favour	8
Against	20
Undecided	32

Testing Independence of Variables:- More often, we ask questions concerning the interrelationships between variables. For example, we ask

1. Is there a difference in the crime rate between children of poor families and those of elite or rich families?
2. Is there a difference in the prevalence of malnutrition between the rural children and urban children?

All the questions raised above have same common characteristics:

- They deal with two or more nominal or ordinal categories
- The categories are non-overlapping
- The data consist of a frequency count
- The data can be cross-classified to fall into several categories of the row and column variables

Let us designate the two attributes as *A* and *B* where, attribute *A* is assumed to have *r* categories and attribute *B* is assumed to have *c* categories. Furthermore, assume the total number of observations in the problem is *N*. A representation of these observations is shown below in a table where O_{ij} represents the observations in the ith row and jth column. Such a table in the matrix form is called a contingency table and is shown below:

		B							Total
		B_1	B_2	....	B_j	....	B_c		
	A_1	O_{11}	O_{12}	....	O_{1j}	....	O_{1c}		R_1
	A_2	O_{21}	O_{22}	....	O_{2j}	....	O_{2c}		R_2
A	$\vdots$								
	A_i	O_{i1}	O_{i2}	....	O_{ij}	....	O_{ic}		R_i
	$\vdots$								
	A_r	O_{r1}	O_{r2}	....	O_{rj}	....	O_{rc}		R_r
Total		C_1	C_2	....	C_r				

In the table, R_i is the total of ith row and C_j is the total of jth column. The frequencies in these cells are termed as cell frequencies and the totals of the frequencies in each of the rows (R_i) and columns (C_j) are termed as marginal frequencies.

In the analysis of contingency table, the objective is to determine whether or not one method of classification is 'contingent' or 'dependent' on the other method of classification. If not, then the two methods of classification are said to be 'independent'. The null hypothesis to be tested in the contingency table is that *A* and *B* are independent, that is there is no association or relationship between the two variable.

If two criteria of classification are independent, a joint probability is equal to the product of the two corresponding marginal probabilities. Thus, the expected cell frequencies are given by the formula:

$$E_{ij} = \frac{R_i}{N} \times \frac{C_j}{N} \times N = \frac{R_i \times C_j}{N}$$

To conduct the test, we use χ^2 statistic

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{O_{ij}^2}{E_{ij}} - N \sim \chi^2_{(r-1)(c-1)}$$

Sign Test: Suppose that paired or matched random samples are taken from a population and the differences equal to 0 are discarded, leaving *n* observations. The sign test can be used to test the null hypothesis that the population median of the differences is 0. Calculate the difference for each pair of observations and record the sign of this difference. The sign test is used to test:

$$H_0 : \pi = 0.5$$

where π is the proportion of non-zero observations in the population that are positive. The test statistic ***S*** for the sign test for paired sample is simply,

$$S = \text{the number of pairs with a positive difference}$$

and ***S*** has a binomial distribution with $\pi = 0.5$ and $n = \text{the number of non} - \text{zero differences}$

The p-value for a sign test is found using the binomial distribution with $n = \text{number of nonzero differences}$, $S = \text{number of positive differences}$, and $\pi = 0.5$

(a) For an upper-tail test, $H_1 : \pi > 0.5$ $p - value = P(x \geq S)$
(b) For a lower-tail test, $H_1 : \pi < 0.5$ $p - value = P(x \leq S)$
(c) For a two-tail test, $H_1 : \pi \neq 0.5$ $2(p - value)$

Example: An Italian restaurant, close to a college campus, contemplated a new recipe for the sauce used in its pizza. A random sample of eight students was chosen, and each was asked to rate on a scale from 1 to 10 the taste of the original and the proposed new product. The scores of the taste comparisons are shown in the next table, with higher numbers indicating a greater liking of the product. Do the data indicate an overall tendency to prefer the new pizza sauce to the original pizza sauce?

TASTER	RATING		DIFFERENCE	SIGN OF DIFFERENCE
	Original	*New*	*(Original-New)*	
A	6	8	-2	-
B	4	9	-5	-

C	5	4	1	+
D	8	7	1	+
E	3	9	-6	-
F	6	9	-3	-
G	7	7	0	0
H	5	9	-4	-

Solution: Sample size $n = 7$, test-statistic $S = 2$

Wilcoxon Signed Rank Test: One disadvantage of the sign test is that it takes into account only a very limited amount of information- namely the differences. The Wilcoxon Signed Rank Test provides a method to incorporate information about the magnitude of the differences between matched pairs.

The Wilcoxon Signed Rank Test can be employed when a random sample of matched pairs of observations is available. Assume that the population distribution of the differences in these paired samples is symmetric, and we want to test the null hypothesis that this distribution is centered at 0. Discarding pairs for which the difference is 0, we rank the remaining *n* absolute differences in ascending order with ties assigned the average of the ranks they occupy. The sums of the ranks corresponding to positive and negative differences are calculated, and the smaller of these sums is the Wilcoxon Signed Rank Statistic *T*, that is

$$T = \min(T_+, T_-)$$

where $T_+ = \textit{the sum of the positive ranks}$

$T_- = \textit{the sum of the negativs ranks}$

$n = \textit{the number of non-zero difference}$

Example: An Italian restaurant, close to a college campus, contemplated a new recipe for the sauce used in its pizza. A random sample of eight students was chosen, and each was asked to rate on a scale from 1 to 10 the taste of the original and the proposed new product. The scores of the taste comparisons are shown in the next table, with higher numbers indicating a greater liking of the product. Do the data indicate an overall tendency to prefer the new pizza sauce to the original pizza sauce?

TASTER	RATING		DIFFERENCE	RANK (+)	RANK (-)
	Original	*New*	*(Original-New)*		
A	6	8	-2		3
B	4	9	-5		6
C	5	4	1	1.5	
D	8	7	1	1.5	
E	3	9	-6		7
F	6	9	-3		4
G	7	7	0		
H	5	9	-4		5
				Sums 3	Sums 25
Wilcoxon Signed Rank Statistic $T = \min(3,25) = 3$					

Under the null hypothesis that the population differences are centered on 0, the Wilcoxon Singed Rank Test has mean and variance given by

$$E(T) = \mu_T = \frac{n(n+1)}{4},\ Var(T) = \sigma_T^2 = \frac{n(n+1)(2n+1)}{24}$$

Then, for large n, the distribution of the random variable, Z, is approximately standard normal where

$$Z = \frac{T - \mu_T}{\sigma_T}$$

If the number n of non-zero differences is large and T is the observed value of the Wilcoxon statistic, then the following tests have significance level α

(a) For a one-sided lower-tailed alternative hypothesis, the decision rule is:

$$\text{Reject } H_0 \text{ if } Z = \frac{T - \mu_T}{\sigma_T} < -Z_\alpha$$

(b) For a one-sided upper-tailed alternative hypothesis, the decision rule is:

$$\text{Reject } H_0 \text{ if } Z = \frac{T - \mu_T}{\sigma_T} > Z_\alpha$$

(c) For a two-sided alternative hypothesis, the decision rule is:

$$\text{Reject } H_0 \text{ if } Z = \frac{T - \mu_T}{\sigma_T} < -Z_{\alpha/2} \text{ or Reject } H_0 \text{ if } Z = \frac{T - \mu_T}{\sigma_T} > Z_{\alpha/2}$$

Mann-Whitney U Test (Independent Samples Test): Assume that apart from any possible differences in central location, that the two population distributions are identical. Suppose that n_1 observations are available from the first population and n_2 observations from the second. The two samples are pooled and the observations are ranked in ascending order, with ties assigned the average of the next available ranks. Let R_1 denote the sum of the ranks of the observations from the first population. The **Mann-Whitney U statistic** is then defined as

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

Assuming that the null hypothesis that the central locations of the two population distributions are the same, the **Mann-Whitney U,** has mean and variance

$$E(U) = \mu_U = \frac{n_1 n_2}{2}, \quad Var(U) = \sigma_U^2 = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$$

Then for large sample sizes (both at least 10), the distribution of the random variable

$$Z = \frac{U - \mu_U}{\sigma_U}$$

is approximated by the normal distribution.

Decision Rule: It is assumed that the two population distributions are identical, apart from any possible differences in central location. In testing the null hypothesis that the two population distributions have the same central location, the decision rule for a given significance level is

(a) For a one-sided lower-tailed alternative hypothesis, the decision rule is:

$$\text{Reject } H_0 \text{ if } Z = \frac{U - \mu_U}{\sigma_U} < -Z_\alpha$$

(b) For a one-sided upper-tailed alternative hypothesis, the decision rule is:

$$\text{Reject } H_0 \text{ if } Z = \frac{U - \mu_U}{\sigma_U} > Z_\alpha$$

(c) For a two-sided alternative hypothesis, the decision rule is:

$$\text{Reject } H_0 \text{ if } Z = \frac{U - \mu_U}{\sigma_U} < -Z_{\alpha/2} \text{ or Reject } H_0 \text{ if } Z = \frac{U - \mu_U}{\sigma_U} > Z_{\alpha/2}$$

Example: Do the data indicate a difference in the median number of hours per week that students spend studying for introductory finance and accounting courses?

Finance	10	6	8	10	12	13	11	9	5	11		
Accounting	13	17	14	12	10	9	15	16	11	8	9	7

Wilcoxon Rank Sum Test (Independent Samples Test): The **Wilcoxon Rank Sum Test** is similar to the **Mann-Whitney U Test**. The results will be the same for both tests.

Suppose that n_1 observations are available from the first population and n_2 observations from the second. The two samples are pooled and the observations are ranked in ascending order, with ties assigned the average of the next available ranks. Let T denote the sum of the ranks of the observations from the first population (T in the Wilcoxon Rank Sum Test is the same as R_1 in the Mann-Whitney U Test). Assuming that the null hypothesis to be true, the **Wilcoxon Rank Sum Statistic** T has

$$E(T) = \mu_T = \frac{n_1(n_1 + n_2 + 1)}{2} \text{ and } Var(T) = \sigma_T^2 = \frac{n_1 n_2(n_1 + n_2 + 1)}{12}$$

Then for large samples (both at least 10), the distribution of the random variable

$$Z = \frac{T - \mu_T}{\sigma_T}$$

is approximated by the normal distribution.