NON-PARAMETRIC TESTS

7a

Many common statistical tests like t, Z and F, used to test the hypothesis concerning population parameters like, population mean (μ), population variance$(\sigma)^2$ etc., are based on certain assumptions about the population or the parameters of the population and hence they are called parametric tests. The most important assumption in the application of these tests is that the populations from which samples are drawn follow normal distribution. Also the usual t-test for testing the difference of two means requires that samples be drawn from populations which are normally distributed and should have equal variances. For small samples, if the population is not normally distributed, the conclusion based on parametric tests will no longer be valid. For such situations, statisticians have devised alternate procedures, which do not require the above assumptions about the population distribution (except some mild or weaker assumptions such as observations are independent and continuous are to be satisfied) and are known as non-parametric or distribution free tests. These tests are applicable even when the data are measured on nominal and ordinal scales and utilize some simple aspects of sample data such as sign of measurements or rank statistic or category frequencies. 7c

Advantages of Non-Parametric Tests:

7b

i) No assumption is made about the distribution of the parent population from which the sample is drawn.

ii) These tests are readily comprehensible and easy to apply.

iii) Since the data in social sciences, in general, do not follow normal distribution, hence non-parametric tests are frequently applied by social scientists.

iv) Non-parametric methods are applicable to data which are given in ranks or grades like A^+, A, B, B^-, C etc.

v) If the sample size is as small as 6, there is no alternative except to use a non-parametric test unless the nature of the population distribution is exactly known.

Disadvantages:

i) Non parametric tests can result in loss of much of the information contained within the data.

ii) They are less efficient and less powerful than their counter parametric tests.

iii) The drawback of ranking, on which many non-parametric tests are based, is that the magnitude of the observations is lost and so the important information within the data is not properly utilized.

From the above discussion, it is clear that whenever the assumption of parametric tests are met, then the non-parametric tests should not be used, as these tests are not as sensitive and powerful as the classical parametric tests, in dealing with the data. When we are unable to apply parametric tests only then we resort to non-parametric tests. In this chapter, various non-parametric tests applicable under different situations are discussed and illustrated with examples.

7.1 One Sample Tests:

These tests lead us to decide whether the population follows a known distribution or the sample has come from a particular population. We can also test whether the median of the population is equal to a known value. A test is also given to test the randomness of a sample drawn from a population as it is the crucial assumption in most of the testing procedures.

7.1.1 Runs Test for Randomness:

7d

One of the most important aspects of all types of statistical testing is that the sample selected must be representative of the population as far as possible since decisions about the population are to be made on the basis of results obtained from samples. Thus the requirement of the randomness of the sample is mandatory. However, in many situations, it is difficult to decide whether the assumption of randomness is fulfilled or not. The assumption of randomness can be tested by runs test, which is based on the theory of runs.

Run and Run Length: A run is defined as a sequence of identical letters (symbols) which are followed and preceded by different letters or no letter at all and number of letters in a run is called run length.

Suppose that after tossing a coin say 20 times, following sequence of heads (H) and tails (T) occur:

HH	TTT	HHH	T	HH	TT	HH	T	HH	TT
1	2	3	4	5	6	7	8	9	10

The first sequence of HH is considered as a run of length 2. Similarly occurrence of TTT is considered as another run of length 3 and so on. So counting the runs in similar way, the total number of runs occurred in above sequence is 10 i.e. R = 10.

The total number of runs (R) in any given sample indicates whether the sample is random are not. Too many or too small number of runs creates doubt about the randomness of sample.

Test Procedure:

Generalizing the problem, let one kind of elements be denoted by plus (+) sign and second kind of elements be denoted by minus (-) sign. The concept of ó+ô and ó-ó provides the direction of change from an established pattern. In the above example if óHô is denoted by ó+ô and óTô denoted by ó-ó then we have

n_1 = number of ó+ô signs = 11

n_2 = number of ó-ô signs = 9

and total sample size $n = n_1 + n_2 = 20$

while the number of runs (R) = 10

Now if sample size is small so that n_1 and n_2 are less than or equal to 20 each, then to test the hypothesis:

H_0: Observations in the sample are random against

H_1: Observations in the sample are not random

Compare the observed number of runs (R) in the sample with two critical values of (R) for given values of n_1 and n_2 at a predetermined level of significance (critical values of óRô for Runs test are available in the Appendix of various statistics books)

Decision Rule:

Reject H_0 if $R \leq c_1$ or $R \geq c_2$, otherwise accept H_0 where c_1 and c_2, are two critical values and may be obtained from standard tables.

For the above example the critical value c_1 from less than table for $n_1 = 11$ and $n_2 = 9$ is ó6ô and from more than table the critical value c_2 is 16. So our acceptance region is

$6 < r < 16$. Since the observed value of "R" is 10, so H_0 is accepted, hence we conclude that our sample observations are drawn at random.

Large sample runs test:

If the sample size is large so that n_1 or n_2 is more than 20, the sampling distribution of "R" can be closely approximated by the normal distribution with mean and variance.

$$E(R) = \varepsilon_r = \frac{2n_1n_2}{n_1 + n_2} + 1$$

$$V(R) = \sigma^2_r = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}$$

Thus $Z = \dfrac{R - \varepsilon_r}{\sigma_r} \sim N(0,1)$

So, if $|Z_{cal}|$ $\geq$ 1.96 then reject H_0 at 5% level of significance and conclude that the sample is not random

$<$ 1.96 then accept H_0 at 5% level of significance

7g

Example-1: A researcher wants to know whether there is any pattern in arrival at the entrance of the shopping mall in terms of males and females or simply such arrivals are random. One day he stationed himself at the main entrance and recorded the arrival of Men (M) and Women (W) of first 40 shoppers and noted the following sequence.

M WW MMM W MM W M W M WWW MMM W MM WWW MMMMMM WWW MMMMMM

Test randomness at 5% level of significance

Solution: H_0 : Arrival of men and women is random

H_1 : Arrival is not random

Here Number of men $(n_1) = 25$

Number of women $(n_2) = 15$

Number of runs $(R) = 17$

Since here $n_1 = 25$ is > 20 so sampling distribution of "R" is approximated by normal distribution with mean:

$$\varepsilon_r = \frac{2n_1n_2}{n_1 + n_2} + 1 = \frac{2 \text{ x } 25 \text{ x } 15}{25 + 15} + 1 = \frac{750}{40} + 1 = 19.75$$

$$\text{and } \mathring{u}_r = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}}$$

$$= \sqrt{\frac{2 \text{ x } 25\text{x } 15\,(2 \text{ x } 25 \text{ x } 15 - 25 - 15}{(25 + 15)^2\ (25 + 15 - 1)}}$$

$$\mathring{u}_r = \sqrt{\frac{750 \text{ x } 710}{1600 \text{ x } 39}} = 2.39$$

$$\text{Thus, } Z_{cal} = \frac{R - \varepsilon_r}{\mathring{u}_r} = \frac{17 - 19.75}{2.39} = -1.15$$

$\therefore$ Since $|Z_{cal}| < 1.96$ so H_0 is accepted i.e. sample is considered as random.

Note: To test the randomness of a sample of n observations, runs can also be generated by considering positive and negative signs of the deviations of observations from the median of the sample.

Let x_1, x_2, é .., x_n be a sample of size n drawn in that order and we wish to test:

H_0 : Observations in the sample are drawn at random

H_1 : Observations in the sample are not random

Let M be the sample median. For generating runs of positive and negative signs, we compute the deviations x_1 ï M, x_2 ï M,·····,x_n ï M and consider only the signs of these deviations.

Let n_1 and n_2 be the number of +ve and -ve signs so that $n_1 + n_2 \leq n$ (ignoring the zero deviation) and R be the total number of runs of +ve and ï ve signs in the sample.

Decision Rule:

Follow the usual procedure of runs test as given above.

Example-2: Following measurements were recorded in a sample of 20 earheads in a

78 Wheat variety V_1: 8.9, 8.4, 10.3, 11.1, 7.8, 9.3, 9.9, 8.2, 10.9, 10.3, 10.8, 8.6, 9.4, 8.9, 9.4, 8.9, 9.5, 9.9, 9.6, 9.7, 9.2 and 10.0. Test the randomness of the sample using runs test.

Solution:

H_0 : Sample is drawn at random

H_1 : Sample is not drawn at random

Let Ŭ= 0.05

Test Statistic: Median is found to be 9.55, therefore, generating runs of the +ve and ï ve signs, we have.

8.9	8.4	10.3	11.1	7.8	9.3	9.9	8.2	10.9	10.3
-	–	+	+	–	–	+	–	+	+
10.8	8.6	9.4	8.9	9.5	9.9	9.6	9.7	9.2	10.0
+	–	–	–	–	+	+	+	–	+

Find the median which is the arithmetic mean of two middle observations 9.5 and 9.6 come out to be equal to 9.55

Consider the signs of the deviation x_i ï 9.55, i = 1, 2,·····,20 and count the number of runs R as the test statistic:

- - + + - - + - + + + - - - - + + + - +

Here number of plus signs (n_1) = 10

Number of minus signs (n_2) = 10

Number of runs (R) = 10

Conclusion:

Critical value c_1 for n_1 = 10 and n_2 = 10 (at Ŭ= 0.05) = 6

Critical value c_2 for n_1 = 10 and n_2 = 10 (at Ŭ= 0.05) = 16

The test statistic R (=10) lies in the acceptance region 6 < R < 16. Hence H_0 is not rejected and we conclude that the sample is considered as drawn at random.

Non-Parametric Alternative to One Sample t-test:

7.1.2 Sign Test:

72

The sign test is used for testing the median rather than mean as location parameter of a population i.e. whether the sample has been drawn from a population with the specified median M_o. It is a substitute of one sample t-test when the normality assumption of the parent population is not satisfied. This test is the simplest of all the non-parametric tests and its name comes from the fact that it is based upon the signs of the differences and not on their numerical magnitude.

Let x_1, x_2,·····,x_n be a random sample of size ónô from a population with unknown median M.

It is required to test the null hypothesis H_0: M = M_0 (some pre-specified value) against a two-sided (H_1: M ≠ M_0) or one sided (H^*_1: M > M_0 or H_1^{**} : M < M_0) alternative.

If the sample comes from the distribution with median M_0, then on the average, half of the observations will be greater than M_0 and half will be smaller than Mo. Compute the differences (x_i - M_0) for i = 1, 2,·····,n and consider their signs and ignore the sample values equal to M_0. Let the number of plus and minus signs be r and s respectively, with r + s ≤ n. For test statistic, we consider only the plus signs (r).

The distribution of ŕrô for given n is a binomial distribution with p = P (X > M_0) = 0.5. Thus the null hypothesis becomes:

H_0: p = 0.5 vs H_1: p ≠ 0.5 or H_1^* : p > 0.5 or H_1^{**} : p < 0.5. In case there is pre-information that sample has an excess number of plus signs we may use one tailed test (i.e. H_1^* : p > 0.5). Similarly, alternative hypothesis H_1^{**} : p < 0.5 will be chosen when it is expected that the sample will have few plus signs. The only difference between one tailed and two-tailed tests is that of critical values for a prefixed α.

Test criterion:

For a two tailed test, reject H_0 if r ≥ $r_{\alpha/2}$ or ≤ $r'_{\alpha/2}$ where $r_{\alpha/2}$ and $r'_{\alpha/2}$ are the critical values at the significance level α. Here $r_{\alpha/2}$ is the smallest integer such that

$$\sum_{r=r_{\alpha/2}}^{n} \binom{n}{r} \left(\frac{1}{2}\right)^n \leq \alpha / 2$$

and $r'_{\alpha/2}$ is the largest integer such that

$$\sum_{r=0}^{r'_{\alpha/2}} \binom{n}{r} \left(\frac{1}{2}\right)^n \leq \alpha / 2$$

for one tailed alternative $H_1^* - p > 1/2$, reject H_0 if r≥ r_α where r_α is the smallest integer such that

$$\sum_{r=r_{Ŭ}}^{n} \binom{n}{r} \left(\frac{1}{2}\right)^n \leq Ŭ$$

and for $H_1^{**} - p > 1/2$, reject H0 if r ≤ r'_α where r'_α is the largest integer such that

$$\sum_{r=0}^{r'_{\alpha}} \binom{n}{r} \left(\frac{1}{2}\right)^{n} \leq \alpha$$

Large Sample Approximation:

If r + s > 25, then normal approximation to the binomial may be used. In that case test statistic Z is computed as under:

$$Z = \frac{r - (r+s)/2}{\sqrt{(r+s)/4}} = \frac{r-s}{\sqrt{r+s}}$$

If $|Z_{cal}| \geq Z_{tab}$ at a specified Ŭ then we reject H_0.

7g

Example-3: Following measurements were recorded for length (cms) of 20 randomly selected earheads of a wheat variety V. 8.9, 8.4, 10.3, 11.1, 7.8, 9.3, 9.9, 8.2, 10.9, 10.3, 10.8, 8.6, 9.4, 8.9, 9.5, 9.9, 9.6, 9.7, 9.2 and 10.0. Test the hypothesis that the median length of earheads is (i) equal to 9.5 cm (ii) more than 9.5 cm (iii) less than 9.5 cm at α = 0.05.

Solution: Obtain the signs of the differences (x_i ï 9.5), which are pasted below the measurements. We find that there are 10 plus, 9 minus and one zero differences. Ignore the zero difference so that the sample size n reduced to 19. In this case n = 19 and r (number of plus signs = 10).

8.9	8.4	10.3	11.1	7.8	9.3	9.9	8.2	10.9	10.3
-	−	+	+	−	−	+	−	+	+
10.8	8.6	9.4	8.9	9.5	9.9	9.6	9.7	9.2	10.0
+	−	−	−	Ignored	+	+	+	−	+

Conclusion:

i) The critical region for a two sided alternative (H_1 : M Í 9.5 cms) at α = 0.05 is given by $r \geq r_{\alpha/2} = 14$ or $r \leq r_{\alpha/2} = 4$ Since here r = 10 which is neither greater than equal to 14 nor less than equal to 4 thus the null hypothesis is not rejected. Thus we conclude that the median length of ear heads is not significantly different from 9.5 cm.

ii) The critical region for a right alternative $(H_1^* : M > 9.5\,\text{cms})$ is given by $r \geq r_\alpha$. We note from the table $r_{.05} = 14$, hence H_0 is not rejected.

iii) Similarly the critical region for left alternative $(H_1^* : M > 9.5\,\text{cms})$ is given by $r \leq r_\alpha^*$. We note from the table $r'_{0.05} = 5,$ hence H_0 is not rejected.

7h **Example-4:** The following data relate to the daily production of cement (in metric tonnes) in a large plant for 30 days.

11.5	10.0	11.2	10.0	12.3	11.1	10.2	9.6	8.7	9.3
9.3	10.7	11.3	10.4	11.4	12.3	11.4	10.2	11.6	9.5
10.8	11.9	12.4	9.6	10.5	11.6	8.3	9.3	10.4	11.5

Use sign test to test the null hypothesis that the plant average daily production (μ) of cements is 11.2 metric tonnes against alternative hypothesis $\mu < 11.2$ metric tonnes at the 0.05 level of significance.

Solution: H_0: Average production of cement is = 11.2 metric tonnes

H_1: Average production of cement is < 11.2 metric tonnes

Let Ŭ= 0.05

Take the +ve and ï ve signs according as the deviation from 11.2 is +ve or ï ve

+	-	0 (Ignored)	-	+	-	-	-	-	-
-	-	+	-	+	+	+	-	+	-
-	+	+	-	-	+	-	-	-	+

Number of plus signs(r) 11

Number of minus signs (s) 18

Number of zeros 1

So the Sample Size reduced to 29 and r = 11

Substituting the values in the formula:

$$Z = \frac{r - (r+s)/2}{\sqrt{(r+s)/4}} = \frac{r-s}{\sqrt{r+s}} = \frac{11-18}{\sqrt{11+18}} = \frac{-7}{\sqrt{29}} \simeq -1.3$$

Since $|Z_{cal}|$ is > ï $Z_{0.05}$ = -1.645, the null hypothesis is not rejected and we conclude that the plantôs average production of cement is equal to 11.2 m tonnes.

7f

7.1.3 Wilcoxon Signed Ranks Test:

Wilcoxon signed ranks test (Wilcoxon, 1945, 1949) is similar to sign test as it is used to test the same hypothesis about the median of the population. The sign test is based only on the signs of differences but Wilcoxon Signed Rank test takes into consideration not only the signs of differences such as positive or negative but also the size of the magnitude of these differences. So this test is more sensitive and powerful than the sign test provided the distribution of population is continuous and symmetric.

Let $x_1, x_2, \ldots, x_n$ denote a random sample of size 'n' drawn from a continuous and symmetric population with unknown median M. We are required to test the hypothesis about the median (M) that is:

$H_0 : M = M_0$ against the alternative hypothesis $H_1 : M \neq M_0$

Choose $\alpha = 0.05$

Take the differences of sample values from M_0 i.e. $d_i = x_i - M_0$, $i = 1, 2, \ldots, n$ and ignore the zero differences. Then assign the respective ranks to the absolute differences in ascending order of magnitude (after ignoring the signs of these differences) so that the lowest absolute value of the differences get the rank '1' second lowest value will get rank '2' and so on. For equal values of absolute differences, average value of ranks would be given.

After assigning the ranks to these differences assign the sign of original differences to these ranks. These signed ranks are then separated into positive and negative categories and a sum of ranks of each category is obtained. Let T^+ denote the sum of ranks of the positive d_i's and T^- denote the sum of ranks of negative differences.

Then clearly $T^{+} + T^{-} = \dfrac{n(n+1)}{2}$, n being the number of non-zero d_i's.

For Small Samples:

Let T = Minimum (T^+, T^-) is taken as test statistic. If $T_{cal} \leq$ critical values of T at specified values of n and α, then H_0 is rejected.

For Large Samples:

For n > 25, T is approximately normally distributed under H_0 with $\mu_T = n(n+1)/4$

and $\sigma_T = \sqrt{n(n+1)(2n+1)/24}$ and $Z_{cal} = \dfrac{T - \varepsilon_T}{\sigma_T}$ is to be compared with z_{tab} for testing H_0 against H_1.

7h

Example-5: Following measurements were recorded for length (in cms) of 20 randomly selected ear-heads of a wheat variety. Test the hypothesis whether median length is equal to 9.5 cm.

8.9	8.4	10.3	11.1	7.8	9.3	9.9	8.2	10.9	10.3
10.8	8.6	9.4	8.9	9.5	9.9	9.6	9.7	9.2	10.0

Solution:

H_0 : M = 9.5 cm

H_1 : M Í 9.5 cms

Let Ŭ= 0.05

The signed deviations of the observations from 9.5 cm with the ranks of their absolute values in parenthesis are:

- 0.6 (9.5), -1.1 (14), 0.8 (11.5), 1.6 (18), -1.7 (19), -0.2 (3.5), 0.4 (6.5),

-1.3(15.5), 1.4(17), 0.8 (11.5), 1.3 (15.5), -0.9(13), - 0.1(1.5), - 0.6 (9.5),

0 (ignored), 0.4 (6.5), 0.1 (1.5), 0.2 (3.5), -0.3(5), 0.5(8)

Here T^+ = 99.5 and T^- = 90.5 so that T= 90.5.

The table value T_α (two tailed test) for n=19 (number of non-zero deviations) at α = 0.05 is 46. Since $T_{cal} > T_{Ŭ}$, therefore H_0 is not rejected and we conclude that median length of the earheads is equal to 9.5 cms.

Using large sample approximation:

$$\mu_T = n(n+1)/4 = 19(20)/4 = 95$$

and $$ů_T = \sqrt{n\,(n+1)\,(2n+1)/24} = \sqrt{19\,(20)\,(39)/24} = 24.8$$

$$Z_{cal} = \frac{T - \varepsilon_T}{ů_T} = \frac{90.5 - 95}{24.8} = -0.18$$

Since $|Z_{cal}|$ = 0.18 is less than Z_{tab} (at Ŭ = 0.05) = 1.96, therefore, H_0 is not rejected.

7.2 Two Sample Tests for Dependent Samples (Non-Parametric Alternative to paired t-test):

7.2.1 Paired Sample Sign Test:

This test is a non-parametric alternative to paired t-test and is used for the comparison of two dependent samples. Here it is desired to test the null hypothesis that the two samples are drawn from the populations having the same median i.e.

$H_0 : M_1 = M_2$ vs H_1 : M_1 Í M_2.

The procedure of single sample sign test explained in section 7.1.2 can be applied to paired sample data.

Here we draw a random sample of 'n' pairs and observations $(x_1, y_1), (x_2, y_2) \ldots (x_n, y_n)$ giving 'n' differences

$$d_i = x_i – y_i \quad \text{for } i = 1, 2, \cdots, n$$

It is assumed that the distribution of differences is continuous in the vicinity of its median 'M' i.e. $P[d > M] = P[d < M] = 1/2$.

All the procedure of one-sample sign test will remain valid for the paired sample sign test, with d_i displaying the role of x_i in one sample sign test.

7.2.2 Paired Sample Wilcoxon Signed Ranks Test:

This test (Wilcoxon, 1945, 1949) deals with the same problem as the paired sample sign test and is an extension of one sample Wilcoxon signed ranks test (Section 7.1.3). But this test is more powerful than paired sample sign test since it takes into account the sign as well as the magnitude of the difference between paired observations.

The observed differences $d_i = x_i - y_i$ are ranked in the increasing order of absolute magnitude and then the ranks are given the signs of the corresponding differences. The null and alternative hypotheses are the same as in paired sample sign test i.e.

$$H_0 : M_1 = M_2 \text{ vs } H_1 : M_1 \neq M_2$$

If H_0 is true, then we expect the sum of the positive ranks to be approximately equal to the absolute value of the sum of negative ranks. The whole procedure of this test is the same as that of one sample Wilcoxon signed ranks test with the only difference that in this test d_i's are given by

$$d_i = x_i – y_i, \quad i = 1, 2, \ldots, n$$

Example-6: The weights of ten men before and after change of diet after six months are given below. Test whether there has been any significant reduction in weight as a result of change of diet at 5% level of significance.

Solution: The difference between before and after weights in the sample alongwith their signed ranks are given below:

Sr. No.	Weight (kgs.)		Difference (d_i)	Rank
	Before (x_i)	After (y_i)		
1	74	61	13	10
2	80	69	11	9
3	69	61	8	5
4	82	72	10	7.5
5	64	71	-7	-4
6	85	79	6	3
7	71	75	-4	-2
8	91	81	10	7.5
9	84	75	9	6
10	64	62	2	1

H_0 : There is no effect of diet in reducing the weight

H_1 : Diet is effective in reducing the weight

Summing up positive and negative ranks we have

$T^+ = 49$, $\quad T^- = 6$ so that T = Minimum (49, 6) = 6

The critical value of the Wilcoxon statistic (T_α) at $\alpha = 0.05$ (one tailed test) for n = 10 is equal to ($T_{0.05}$) = 10. Since $T_{cal} < T_\alpha$, therefore, H_0 is rejected. Thus we conclude that there is significant reduction in weight as a result of change of diet.

Zi **Example-7:** Two types of package programmes were offered to 30 farmers in an investigation and were used to award scores for each programme on its merits and scores are given below. Test whether there is any significant difference between two types of programmes at Ŭ= 0.05 by (i) Paired sample Wilcoxon signed ranks test (ii) Paired Sign test.

Solution: The differences in the sample values alongwith their signed ranks are given below:

Farmer Sr. No.	Type-I (x_i)	Type II (y_i)	$d_i = x_i - y_i$	Rank	Farmer Sr. No.	Type I (x_i)	Type II (y_i)	$di = x_i - y_i$	Rank
1	64	68	-4	-12	16	39	50	-11	-25
2	70	72	-2	-4.5	17	47	40	7	19
3	65	60	5	14.5	18	35	35	0	ignored
4	72	69	3	8.5	19	57	50	7	19
5	35	42	-7	-19	20	68	52	16	28
6	52	49	3	8.5	21	70	68	2	4.5
7	45	45	0	ignored	22	43	51	-8	-21.5
8	76	73	3	8.5	23	59	55	4	12
9	60	58	2	4.5	24	38	39	-1	-1.5
10	48	54	-6	-16.5	25	59	51	8	21.5
11	39	42	-3	-8.5	26	45	50	-5	-14.5
12	67	54	13	26	27	62	68	-6	-16.5
13	50	65	-15	-27	28	72	63	9	23
14	76	75	1	1.5	29	78	68	10	24
15	42	40	2	4.5	30	48	52	-4	-12

Solution: Here, H_0 : Two package programmes are equally meritorious

H_1 : Two package programmes are not equally meritorious

Counting the sum of positive ranks (T^+) and negative ranks (T^-), we get

$T^+ = 227.5$, $T^- = 178.5$

$T =$ Minimum $(T^+, T^-) = 178.5$

Since the sample size is large, therefore, the normal approximation is used. Hence we have

$$Z_{cal} = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} = \frac{178.5 - \frac{30(31)}{4}}{\sqrt{\frac{30(31)(61)}{24}}} = -1.11$$

Since $|Z_{cal}| = 1.11 < 1.96$, we, therefore, do not reject H_0 and conclude that two package programmes are equally meritorious i.e. there is no significant difference between the merits of two programmes.

ii) Now we will apply paired sign test to solve the above problem. We note that

Number of plus signs (r) = 16

Number of minus signs (s) = 12

Number of zeros = 2, hence sample size reduced to 28.

Since r + s = 28, therefore, normal approximation to the binomial distribution may be used. Hence the value of test statistic is given by

$$Z_{cal} = \frac{r - s}{\sqrt{r + s}} \qquad = \frac{16 - 12}{\sqrt{16 + 12}} = 0.76$$

Since $|Z_{cal}| < 1.96$, therefore, we do not reject H_0.

7.3 Two Sample Tests for the Independent Samples:

Non-Parametric Alternative to two sample t-test:

7.3.1 Mann-Whitney U-test:

This test was developed by Mann and Whitney (1974) and is a non-parametric alternative to the usual two-sample t-test. To apply this test, the two samples that are to be compared should be independent and the variable under consideration should have a continuous distribution. The null hypothesis under this test is that the two population distributions from which the samples have been drawn are identical. Under experimental situation, the null hypothesis may be that two treatments are identical i.e. two treatment effects do not differ significantly.

Procedure: Let the two independent samples $x_1, x_2, \ldots., x_{n1}$ and $y_1, y_2, \ldots., y_{n2}$ be drawn from two populations having distribution functions F(x) and F(y) which are assumed to be continuous. The hypothesis to be tested is

$$H_0 : F(x) = F(y) \qquad \text{vs} \qquad H_1 : F(x) \neq F(y)$$

For small samples:

When n_1 and n_2 both are less than 8, the test is described below:

The $(n_1 + n_2)$ observations of both the samples are combined and arranged in ascending order of their magnitude alongwith the identity to the sample of which each observation belong.

Find out U_1 by counting how many scores of Xôs precede (are lower than) each score of Yôs and add them.

Find out U_2 by counting how many scores of Y's precede (are lower than each score of X's and add them.

We reject H_0 for small values of $U = \min(U_1, U_2)$ i.e. if the calculated value of U is less than or equal to the tabulated value for given n_1 and n_2 from the table of critical values for Mann-Whitney U-statistic, then the null hypothesis is rejected.

For moderately large samples:

When size of any one sample lies between 9 and 20, we have following steps:

Combine the $(n_1 + n_2)$ observations belonging to both the samples and rank them into a single group in ascending order of magnitude. For tied score give the average rank. The ranks of the X's (first sample) and Y's (second sample) are summed up as R_1 and R_2 respectively.

A more convenient way of finding U_1 and U_2 is given below:

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

or $\quad U_2 = n_1 n_2 - U_1$

Tables of critical values for moderate sample sizes at a specified α are available and test criterion is the same as explained for small samples.

For large samples:

When one or both the samples are larger then 20, the tables of critical values of U-statistic are not much useful. With the increase in sample size, the sampling distribution of U takes the shape of a normal distribution and the corresponding standard normal variate is given by:

$$Z = \frac{U - n_1 n_2 / 2}{\sqrt{n_1 n_2 (n_1 + n_2 + 1)/12}}$$

Conclusion:

i) For two tailed test if $|Z_{cal}| \geq 1.96$, then H_0 is rejected at $\alpha = 0.05$

ii) For right tailed test if $Z_{cal} \geq 1.64$, then H_0 is rejected at $\alpha = 0.05$

iii) For left tailed test if $Z_{cal} \leq -1.64$, then H_0 is rejected at $\alpha = 0.05$

Example-8: An experiment was conducted for comparing two types of grasses on nine plots of size 5x 2 m each. The yields of grasses per plot (kgs) at the time of harvesting are as given below:

	1	**2**	**3**	**4**	**5**	**6**
Grass I (X)	1.96	2.12	1.64	1.78	1.99	
Grass II (Y)	2.13	2.10	2.24	2.08	2.20	2.32

Test whether:

i) There is significant difference between the two types of grasses in respect of yield;

ii) Is grass I is lower yielder than grass II

Solution: (i) $H_0 : F(x) = F(y)$ i.e. there is no significant difference in the yields of two types of grasses

$H_1 : F(x) \neq F(y)$

Combining the data of both the samples and ranking them in ascending order of magnitude, we have

Yield	1.64	1.78	1.96	1.99	2.08	2.10	2.12	2.13	2.20	2.24	2.32
Rank	1	2	3	4	5	6	7	8	9	10	11
Sample	X	X	X	X	Y	Y	X	Y	Y	Y	Y

Calculating the values of U_1 and U_2 by counting the number of X's preceding (are lower than) each Y and adding up and vice versa.

$U_1 = 4 + 4 + 5 + 5 + 5 + 5 = 28$

and $U_2 = 0 + 0 + 0 + 0 + 2 = 2$

The critical value of U-statistic at $\alpha = 0.05$ ($n_1 = 5$, size of smaller sample and $n_2 = 6$, size of large sample) for a two tailed test = 3

Since $U = \text{Min}(U_1, U_2) = 2$ which is less than critical value, therefore, H_0 is rejected. Thus we conclude that there is significant difference in the yields of two types of grasses.

ii) In this case $H_1 : F(x) < F(y)$

Critical value of U at $\alpha = 0.05$ for one tailed test = 5

Since calculated $U_2 = 2$ is less than the critical value, therefore, H_0 is rejected. Thus it is concluded that grass I is lower yielder than grass-II.

Example-9: A random sample of 15 students was taken from private school in city and another random sample of 12 students was taken from public school in the same city and was administered the same test in English. The scores of students from both schools were recorded as follows:

Sr. No.	School (A)	Ranks	School (B)	Ranks
1	73	14	70	11
2	75	16	78	19
3	83	23.5	79	20
4	77	18	81	22
5	72	13	65	7
6	69	10	63	5
7	56	2	74	15
8	80	21	83	23.5
9	68	9	67	8
10	60	3	76	17
11	84	25	88	27
12	61	4	48	1
13	64	6		Sum (R_2)=176
14	71	12		
15	86	26		
		Sum (R_1) = 202		

Test whether the quality of education in English in private schools is similar to the one in public schools.

Solution: After mixing both the samples and arranging in ascending order of their magnitude and assigning ranks we have

48	1	69	10	78	19
56	2	70	11	79	20
60	3	71	12	80	21
61	4	72	13	81	22
63	5	73	14	83	23
64	6	74	15	83	24
65	7	75	16	84	25
67	8	76	17	86	26
68	9	77	18	88	27

Sum of ranks for private school (R_1) = 202

Sum of ranks for public school (R_2) = 176

$n_1 = 15$ and $n_2 = 12$

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = (12)(15) + \frac{15(15+1)}{2} - 202$$

= 180 + 120 - 202 = 98

$$U_2 = (12)(15) + \frac{12(12+1)}{2} - R_2$$

= 180 + 78 ï 176 = 82

$\therefore$ U = minimum (U_1, U_2) = 82

Critical value of U-statistic for $n_1 = 12$ and $n_2 = 15$ at Ǔ= 0.05 (for two tailed test) = 49. Since $U_{cal} = 82$ is more than the critical value, therefore, H_0 is accepted.

Although the sample sizes n_1= 12 and n_2= 15 are not large, but for the sake of illustration, we solve the example through normal approximation.

Thus assuming sample sizes large (for illustration) we have:

$$E(U) = \frac{12 \text{ x } 15}{2} = 90$$

$$\text{ů}_u = \sqrt{\frac{12 \text{ x } 15 (12+15+1)}{12}} = \sqrt{\frac{180 (28)}{12}} = 20.5$$

$$\therefore Z_{cal} = \frac{U - E(U)}{\text{ů}_u} = \frac{82 - 90}{205} = -0.39$$

Since |z| = 0.39 < 1.96

We accept H_0 i.e. evidence does not suggest that there is any significant difference in the quality of education in English in public and private schools.

7.3.2 Siegel-Tukey Test for Equal Variability:

This test (Siegel and Tukey, 1960) is used to test the null hypothesis whether two samples have been drawn from the two populations having the same variance i.e. whether the two populations have the equal variability. Siegel-Tukey test is the non-parametric alternative to the corresponding parametric F-test. The only assumptions underlying this test are that the populations have continuous distribution and sample sizes are not too small i.e. $n_1 + n_2 > 20$. So our hypothesis is

H_0 : $\sigma_1^2 = \sigma_2^2$ i.e. two populations have equal variance

H_1 : $\sigma_1^2 \neq \sigma_2^2$

Procedure:

Draw random samples of sizes n_1 and n_2 from the two populations. Then the observations of the two samples are combined and arranged in order of their increasing size. Ranks are allocated according to the following scheme:

- The lowest value is ranked 1.
- The highest two values are ranked 2 and 3 (the largest value is given rank 2)
- The lowest two unranked values are ranked 4 and 5 (the smallest value is given the rank 4).
- The highest two unranked values are ranked 6 and 7 (the largest value is given the rank 6)

This procedure continues, working from the end towards the centre, until no more than one unranked value remains. That is to say, if the number of values is odd, the middle value has no rank assigned to it.

Let n_1 and n_2 denote the sizes of the two samples and let $n_1 \leq n_2$. Let R_x and R_y denote the rank sums of two series X and Y and let R be the minimum of the rank sums of two series. Then the value of test statistic is given by

$$Z = \frac{R - n_1(n_1 + n_2 + 1)/2 + \frac{1}{2}}{\sqrt{n_1 n_2 \,(n_1 + n_2 + 1/12}} \simeq N\,(0,\,1)$$

If $|Z_{cal}| > 1.96$ reject H_0 at Ŭ= 0.05

and if $|Z_{cal}|$ Ò1.96 accept H_0

Example-10: An agronomist wants to know if a particular variety of paddy gives the same variability or spread in the yield under two different agronomic practices X and Y. Thus he laid the experiment in 10 plots for each practice and the yields (q ha^{-1}) are given below:

x	:	57.3	56.1	52.4	58.8	58.5	60.1	60.1	59.8	62.6	59.4
y	:	63.1	52.9	53.6	65.3	66.5	53.6	54.2	61.7	57.3	54.9

Apply Siegal-Tukey rank sum dispersion test to test the assertion.

Solution: Rank assignment to the combined data of two samples X and Y

Sample **Value** **Rank**	x 52.4 1	y 52.9 4	y 53.6 5	y 53.6 8	y 54.2 9	y 54.9 12	x 56.1 13	x 57.3 16	y 57.3 17	x 58.5 20
Sample **Value** **Rank**	x 58.8 19	x 59.4 18	x 59.8 15	x 60.1 14	x 60.1 11	y 61.7 10	x 62.6 7	y 63.1 6	y 65.3 3	y 66.5 2

Here $n_1 = n_2 = 10$

$R_x = 1 + 13 + 16 + 20 + 19 + 18 + 15 + 14 + 11 + 7 = 134$

$R_y = 4 + 5 + 8 + 9 + 12 + 17 + 10 + 6 + 3 + 2 = 76$

Hence R Minimum (134, 76) = 76

$$Z = \frac{76 - 10\,(10+10+1)/2 + \frac{1}{2}}{\sqrt{(10\,\mathrm{x}10)\,(10+10+1)/12}} = \frac{-28.5}{\sqrt{175}} = \frac{-28.5}{13.23} = -2.15$$

Here $|Z_{cal}| = 2.15 > 1.96$, thus we reject the null hypothesis and conclude that there is significant difference in variability in the yields of paddy under two different practices.

7.4 k-Sample Tests ($k \geq 3$)

1. Median Test:

The median test is used to test the null hypothesis whether two or more populations have the same median or not. To test this assertion, random samples of sizes $n_1, n_2, \cdots, n_k$ are drawn from ókô populations. The observations obtained in each sample are combined and median of the combined sample is determined. This is called grand median say (M). Let O_{1i} ($i = 1, 2, \cdots, k$) denote the number of observations in the i^{th} sample that exceed the grand median ómô and let O_{2i} denote the number of observations in the i^{th} sample, which are less than or equal to the grand median. Arrange these frequencies into 2 x k contingency table as follows:

	1	**2**	**3**	**…………….**	**k**	**Total**
> Median	O_{11}	O_{12}	O_{13}	é é é é é .	O_{1k}	N_1
≤ Median	O_{21}	O_{22}	O_{23}	é é é é ..é	O_{2k}	N_2
	n_1	n_2	n_3	é é é é é ..	n_k	N

Thus N_1 denotes the number of observations above the grand median and N_2 the number of observations less than or equal to the grand median in all samples then $N = N_1 + N_2$ are the total number of observations.

Assumptions:

1. Samples are drawn randomly and independent of each other.
2. The measurement scale is ordinal

Test Procedure:

H_0 : Median of all ‘k’ populations is same, against the alternative.

H_1 : At least two populations have different median.

The rest of the procedure is the same as that of chi-square test for independence.

Test statistic: Let E_{ij} denote expected frequency of $(i, j)^{th}$ cell in 2 x k contingency table (i = 1, 2 and j = 1, 2,·····,k). Now under null hypothesis, the expected frequencies are given by:

$$E_{1i} = \frac{N_1}{N} n_i \text{ and } E_{2i} = \frac{N_2}{N} n_i$$

Thus test statistic for 2 k categories is given by:

$$T = \sum_{i=1}^{2} \sum_{j=1}^{k} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

which approximately follows a chi-square distribution with ‘k-1’ degree of freedom

Test Criteria:

If the value of $\left(\frac{N-1}{N}\right)T$ is greater than or equal to tabulated value of χ^2 for (k-1) d.f. reject H_0, otherwise H_0 is not rejected.

Note: For k = 2, i.e. for the comparison of two populations the contingency table is given below:

	Sample –I	**Sample – II**	**Total**
> Median	a	b	R_1
< Median	c	d	R_2
Total	C_1	C_2	

$$G^2_{cal} = \frac{N(ad - bc)^2}{R_1 R_2 C_1 C_2}$$

If N < 50 or any cell frequency is less than 5, then Yatesô correction of continuity is applied and we get:

$$G^2_{cal} = \frac{N[|ad - bc| - N/2]^2}{R_1 R_2 C_1 C_2}$$

If $G^2_{cal} \geq G^2_{1,\alpha}$, the critical value of χ^2 distribution with 1 df at a given α, then reject H_0.

Example-11: The scores of two groups of students are given below:

Group –I: 61, 64, 59, 53, 56, 52, 57, 54, 51, 57, 60, 62

Group-II: 60, 58, 57, 64, 67, 70, 63, 62, 52, 53, 55, 56

Test by applying median test if the two groups differ significantly

Solution: After arranging the combined data of two groups in ascending order of magnitude, we get:

51, 52, 52, 53, 53, 54, 55, 56, 56, 57, 57, 57, 58, 59, 60, 60, 61, 62, 62, 63, 64, 64, 67, 70

Combined median = Average 12^{th} and 13^{th} terms in the ordered data = (57 + 58)/2 = 57.5

The values above and below the median in the two groups are determined and put in 2 x 2 table as:

	Group-I	Group-II	Total
No. of scores > Median	5	7	12
No. of Scores < Median	7	5	12
	12	12	24

$$G^2_{cal} = \frac{N(ad - bc)^2}{R_1 R_2 C_1 C_2} = \frac{24[25 - 49]^2}{12 \text{ x } 12 \text{ x } 12 \text{ x } 12} = \frac{24 \text{ x } 24 \text{ x } 24}{12 \text{ x } 12 \text{ x } 12 \text{ x } 12} = 0.67$$

Since G^2_{cal} is less than tabulated value of G^2 distribution with 1 d.f. at 5% level (3.84), therefore, H_0 is not rejected and we conclude that the samples are drawn from the populations with the same median.

gd

2. Kruskal-Wallis H-Test:

To test whether several independent samples have come from identical populations, analysis of variance is the usual procedure provided the assumptions underlying are fulfilled. But if the assumptions are not fulfilled then Kruskal– Wallis test (Kruskal and Wallis, 1952) is used for one way classification data (i.e. completely randomized design). It is an extension of Mann Whitney U-test in which only two independent samples are considered.

Kruskal Wallis test is an improvement over the median test. In the median test the magnitude of various observations was compared with median value only. But in this method the magnitude of observations is compared with every other observation by considering their ranks.

So, if n_1, $n_2,\cdots,n_k$ denote the sample sizes of ‘k’ samples selected from k-populations and $N = n_1+n_2,\cdots,n_k$ denote the total number of observations in ‘k’ samples, then combined sample of ‘N’ observations in increasing order of magnitude will have ranks from 1 to N. Clearly the sum of ranks will be equal to N (N+1)/2. So under the null hypothesis H_0 : All population are identical vs H_1: Populations are not identical.

Let R_1, $R_2,\cdots,R_k$ denote the sum of ranks of the observations of 1st, 2nd, kth sample respectively. Then Kruskal-Wallis H- statistic is given as under:

$$H = \left(\frac{12}{N(N+1)}\right)\left[\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \ldots.. + \frac{R_k^2}{n_k}\right] - 3(N+1)$$

$$= \left(\frac{12}{N(N+1)}\right)\sum_{i=1}^{k}\frac{R_1^2}{n_i} - 3(N+1)$$

Under H_0, the sampling distribution of ‘H’ statistic is approximately chi-square with ‘k-1’ degrees of freedom. Hence reject H_0 if calculated value of H is greater than or equal to critical value of χ^2 with (k-1) df and at a given level ‘α’ otherwise H_0 is not rejected.

Example-12: Sample of three brands of cigarettes were tested for nicotine content. The observations (in mg) are:

Brand 1	7.7, 7.9, 8.5, 8.0, 8.4, 9.1
Brand 2	9.2, 8.6, 9.5, 8.7
Brand 3	8.9, 8.6, 7.8, 10.1, 10.0

Use Kruskal-Wallis Test to test if nicotine contents in three brands are equal.

Solution:

H_0 : The nicotine contents in the three brands are equal

H_1 : The nicotine contents in the three brands are not equal

Here $n_1 = 6$, $n_2 = 4$, $n_3 = 5$ so that $N = 15$

The ranks of corresponding observation are

Brand 1 1, 3, 6, 4, 5, 11

Brand 2 12, 7.5, 13, 9

Brand 3 10, 7.5, 2, 15, 14

Note: Two identical values (8.6) have been given the average rank (7 + 8)/2 = 7.5. The rank totals are $R_1 = 30$, $R_2 = 41.5$ and $R_3 = 48.5$

The value of test statistic H is given by:

$$H_{cal} = \frac{12}{15 \text{ x } 16}\left(\frac{30^2}{6} + \frac{41.5^2}{4} + \frac{48.5^2}{5}\right) - 3 \text{ x } 16$$

$$= \quad 0.05\ (150 + 430.06 + 470.45)\ ï\ \ 48 = 4.56$$

The tabulated value of χ^2 with 2 df at 5% level is 5.99

Since $H_{cal} < \chi^2_{tab}$, therefore, we do not reject H_0 and conclude that nicotine contents in three brands are equal.

Example-13: To test whether four types of treatments differ or not, the treatments were applied to 25 pods randomly. The green pod yield (kg) under four treatments was as under:

Treatments	1	2	3	4	5	6	7
I	3.20	3.35	3.56	2.87	3.89	4.20	3.65
II	3.44	2.88	2.95	3.26	3.98	3.87	3.22
III	3.15	2.66	3.06	2.75	3.45		
IV	2.42	2.30	2.55	2.85	3.05	2.42	

Test the hypothesis that there is no significant difference among the four treatments by (i) Median test (ii) Kruskal Wallis test.

Solution: i) H_0 : Medians of all the treatments are equal.

H_1 : Medians of atleast two treatments are significantly different.

The sequence of combined samples in order of magnitude is

$\underset{1}{[2.30]}$ $\underset{2.5}{[2.42]}$ $\underset{2.5}{[2.42]}$ $\underset{4}{[2.55]}$ $\underset{5}{(2.66)}$ $\underset{6}{(2.75)}$ $\underset{7}{[2.85]}$

$\underset{8}{2.87}$ $\underset{9}{\underline{2.88}}$ $\underset{10}{\underline{2.95}}$ $\underset{11}{[3.05]}$ $\underset{12}{(3.06)}$ $\underset{13}{(3.15)}$ $\underset{14}{3.20}$ $\underset{15}{\underline{3.22}}$

$\underset{16}{\underline{3.26}}$ $\underset{17}{3.35}$ $\underset{18}{\underline{3.44}}$ $\underset{19}{(3.45)}$ $\underset{20}{3.56}$ $\underset{21}{3.65}$ $\underset{22}{\underline{3.87}}$ $\underset{23}{3.89}$

$\underset{24}{\underline{3.98}}$ $\underset{25}{4.20}$

Here k = 4, $n_1 = 7$, $n_2 = 7$, $n_3 = 5$, $n_4 = 6$ and N = 25

Median of the combined sample (M) = 3.15

Treatment	**Observed frequency**		**Total**
	$<$M	**$\geq$ M**	
I	1	6	7
II	2	5	7
III	3	2	5
IV	6	0	6
Total:	**12**	**13**	**25**

Calculation of Expected Frequencies:

Treatment	Category ï I (< M)		Category-II (ÓM)	
	Observed	Expected	Observed	Expected
I	1	$\frac{12}{25}$ x 7 = 3.4	6	$\frac{13}{25}$ x 7 = 3.6
II	2	$\frac{12}{25}$ x 7 = 3.4	5	$\frac{13}{25}$ x 7 = 3.6
III	3	$\frac{12}{25}$ x 5 = 2.4	2	$\frac{13}{25}$ x 5 = 2.6
IV	6	$\frac{12}{25}$ x 6 = 2.8	0	$\frac{13}{25}$ x 6 = 3.2

$$T = \sum_i \sum_j \frac{(Oi_j - E_{ij})^2}{E_{ij}} = \frac{(1-3.4)^2}{3.4} + \frac{(2-3.4)^2}{3.4} + \frac{(3-2.4)^2}{2.4} + \frac{(6-2.8)^2}{2.8} + \frac{(6-3.6)^2}{3.6}$$

$$+ \frac{(5-3.6)^2}{3.6} + \frac{(2-2.6)^2}{2.6} + \frac{(0-3.2)^2}{3.2} = 11.56$$

Test Statistic: $\frac{(N-1)}{N} T = \left(\frac{25-1}{25}\right)(11.56) = 11.10$ (Calculated value)

Since calculated value > 7.815 (tabulated value), therefore, null hypothesis is rejected. Thus median effect of various treatments is significantly different.

ii) The sum of ranks of each individual treatment is

For treatment I R_1 = 8 + 14 + 17 + 20 + 21 + 23 + 25 = 128

For treatment II R_2 = 9 +10 + 15 + 16 + 18 + 22 + 24 = 114

For treatment III R_3 = 5 + 6 + 12 + 13 + 19 = 55

For treatment IV R_4 = 1 + 2.5 + 2.5 + 4 + 7 + 11 = 28

Here $n_1 = 7$, $n_2 = 7$, $n_3 = 5$, $n_4 = 6$

Therefore $N = n_1 + n_2 + n_3 + n_4$

= 7 + 7 + 5 + 6 = 25

Applying Kruskal Wallis test, we have:

$$H = \frac{12}{N(N+1)} \sum \frac{R_i^2}{n_i} - 3(N+1)$$

$$H = \frac{12}{25 \text{ x } 26} \left[\frac{128^2}{7} + \frac{114^2}{7} + \frac{55^2}{5} + \frac{28^2}{6} \right] - 3 \text{ x } 26$$

$$= \frac{12}{25 \text{ x } 26} \text{ x } 4932.81 - 3 \text{ x } 26$$

= 91.067 ï 78 = 13.067

The tabulated value of chi-square $\chi^2_{3,0.05} = 7.815$

Since H_{cal} > 7.815, therefore, null hypothesis is rejected. Thus effect of various treatments is significantly different.

3. Friedman's Test:

This is a non-parametric test for k-related samples of equal size say n, parallel to two-way analysis of variance (Randomized Block Design). Here we have k-related samples of size n arranged in n blocks and k columns in a two-way table as given below:

Blocks	**Samples (Treatments)**							**Block Totals**
	1	2	3	..	..	..	k	
1	R_{11}	R_{12}	R_{13}	..	..	..	R_{1k}	k(k+1)/2
2	R_{21}	R_{22}	R_{23}	..	..	..	R_{2k}	k(k+1)/2
3	R_{31}	R_{32}	R_{33}	..	..	..	R_{3k}	k(k+1)/2
:								
:								
n	R_{n1}	R_{n2}	R_{n3}	..	..	..	R_{nk}	k(k+1)/2
Column totals	$\mathbf{R_1}$	$\mathbf{R_2}$	**..**	**..**	**..**	**..**	$\mathbf{R_k}$	**nk(k+1)/2**

Where R_{ij} is the rank of the observation belonging to sample j in the i^{th} block for j = 1,2,·····,k and i = 1, 2,·····,n. This is the same situation in which there are k treatments and each treatment is replicated n times. Here it should be carefully noted that the observations in a block receive ranks from 1 to k. The smallest observation receives rank 1 at its place and the largest observations receive rank k at its place. Similarly observations receive ranks accordingly in other blocks. Hence the block total are constant and equal to k(k+1)/2, the sum of k integers.

The null hypothesis H_o to be tested is that all the k samples have come from identical populations. In case of experiment design, the null hypothesis H_o is that there is no difference between k treatments. The alternative hypothesis H_1 is that at least two samples (treatments) differ from each other.

Under H_o, the test statistic is:

$$T = \frac{12}{nk\,(k+1)} \sum_{j=1}^{k} R_j^2 - 3n\,(k+1)$$

The statistic T is distributed as χ^2 with (k-1) d.f. Reject H_o if $T_{cal} \geq \chi^2_{\alpha,(k-1)}$, otherwise H_o is not rejected.

Example-14: The iron determination (ppm) in five pea-leaf samples, each under three treatments were as tabulated below:

Sample No. Blocks	Treatments			Block Totals
	1	2	3	
1	591 (1)	682 (2)	727 (3)	6
2	818 (2)	591 (1)	863 (3)	6
3	682 (2)	636 (1)	773 (3)	6
4	499 (1)	625 (2)	909 (3)	6
5	648 (1)	863 (3)	818 (2)	6
Column Totals	**7**	**9**	**14**	**30**

Test the hypothesis that iron content in leaves under three treatments is same.

H_o : That the iron content in leaves under three treatments is the same

H_1: That at least two of them have different effect

Friedman's test statistic is:

$$T_{cal} = \frac{12}{nk\,(k+1)} \sum_{j=1}^{k} R_j^2 - 3n\,(k+1)$$

$$= \frac{12}{5 \times 3 \times 4}(7^2 + 9^2 + 14^2) - 3 \times 5 \times 4$$

= 65.2 ï 60 = 5.2

For Ŭ= 0.05, the table value of $\chi^2_{0.05,\,2} = 5.99$

Since the calculated T-value is less than 5.99, H_o is not rejected. This means that there is no significant difference in the iron content of pea leaves due to treatments.

7.2 Kendall's rank correlation coefficient:

Kendallǿs rank correlation coefficient (Kendall, 1938) τ is a non-parametric measure of correlation and is based upon the ranks of the observations.

Consider each possible pair of individuals (i, j) and the order of this pair in the two rankings. If the pair appears in the same order in both rankings, we allot it a score of +1, and if it appears in reverse orders, a score of -1. The score is thus obtained for each of the $\binom{n}{2} = \frac{n\,(n-1)}{2}$ possible pairs. We then define a rank correlation coefficient τ as

$$\tau = \frac{\text{total score}}{n\,(n-1)/2} \tag{1}$$

Obviously, τ = +1 for perfect agreement, because the score for each pair, each being in the same order in both rankings, is +1 ; and τ = -1 for perfect disagreement, because the score for each pair, being in reverse orders in the two rankings, is now -1.

Suppose the ranking in one series is in the natural order, viz. the order 1, 2,⋯⋯,n. Let us consider the corresponding ranking in the other series. Suppose out of $\binom{n}{2}$ pairs for the second series, P pairs have ranks in the natural order and Q pairs have ranks in the reverse order. Obviously, the P pairs will receive a score of +1 each, while the Q pairs will receive a score of -1 each. Thus, according to (1)

$$Ű= \frac{P-Q}{\binom{n}{2}} = 1 - \frac{2Q}{\binom{n}{2}} = \frac{2P}{\binom{n}{2}} - 1 \text{ since } P + Q = \text{total number of pairs} = \binom{n}{2}$$

Test of Significance:

If the absolute value of τ or S = P ï Q ≥ corresponding value in the table of critical values for Kendall Tau for a given n, then H_0 is rejected.

Example-15: Ten hand writings were ranked by two judges in a competition. The rankings are given below. Calculate Kendallŝ τ coefficient and test for its significance.

Hand-writing										
	A	B	C	D	E	F	G	H	I	J
Judge 1	3	8	5	4	7	10	1	2	6	9
Judge 2	6	4	7	5	10	3	2	1	9	8

Solution: To calculate τ, it is convenient to rearrange one set of ranking so as to put it in the natural order : 1, 2,·····,n. If we do so far the ranking by Judge 1, the corresponding ranking by Judge 2 becomes:

2, 1, 6, 5, 7, 9, 10, 4, 8 and 3

The score obtained by considering the first member 2, in conjunction with the others is 8-1 = 7, because only 1 is smaller than 2. Similarly, the score involving the member 1 is 8, the score involving the member 6 is 4-3=1, and so on. The total score is

S = 7 + 8 + 1 + 2 + 1 ï 2 ï 3 + 0 ï 1= 19-6 = 13

On the other hand, the maximum possible score is (10 x 9)/2 = 45.

Thus τ = 13/45 = 0.289

Test of Significance:

The critical values of τ and S for n = 10 at 5% level of significance are 0.511 and 23 respectively. Since the calculated value of τ and S are less than the critical values, therefore, H_0 is not rejected.

Example-16: Two supervisors ranked 12 workers working under them in order of their efficiency as given below. Calculate Kendallŝ τ coefficient between the two rankings and test for its significance.

Worker	1	2	3	4	5	6	7	8	9	10	11	12
Supervisor (X)	5	6	1	2	3	8.5	8.5	4	7	11	10	12
Supervisor (Y)	5.5	5.5	2	2	2	9	7	4	8	10.5	12	10.5

Solution: We rearrange the ranking of supervisor X in the natural order, and then we have the two sets of ranks as follows:

Supervisor (X)	1	2	3	4	5	6	7	8.5	8.5	10	11	12
Supervisor (Y)	2	2	2	4	5.5	5.5	8	9	7	12	10.5	10.5

The total score in this case is

$$S = 9 + 9 + 9 + 8 + 6 + 6 + 3 + 3 + 3 - 2 - 0 = 54$$

Maximum possible score = n (n-1)/2 = 12(11)/2=66

$$\text{Hence } \tau = \frac{54}{66} = 0.818$$

The critical values of τ and S (for n = 12 at α = 0.05) are 0.455 and 30 respectively. Since calculated values of τ and S are more than critical values, hence H_0 is rejected and we conclude that there is significant correlation between ranking of two supervisors.

Kendall’s Coefficient of Concordance:

Kendall developed coefficient of concordance (K_c) for measuring the relationship between the k variates. It measures the extent of relationship (or the degree of association) between ‘k’ variates based on ‘n’ rankings for each variate. This coefficient avoids the requirement of computing several Spearman rank correlation coefficients pairwise. The two way table of ranks for the variates and the observations are below:

(Ranks)

Variate \ Observation	**1**	**2**	**. . . .**	**n**
1	6	(n – 1)	. . . .	2
2	3	5	. . . .	n-2
.	.	.		.
.	.	.		.
.	.	.		.
k	(n - 5)			1
Total	C_1	C_2		C_n

The following of concordance is given by:

$$K_c = \frac{\sum_{j=1}^{n} C_j^{\ 2} - \frac{(\Sigma C_j)^2}{n}}{\frac{1}{12} k^2 n (n^2 - 1)}$$

where C_j denote the j-th column total for j = 1, 2,·····,n and n be the number of observations in each variate and k be the number of variates. K_c always lies between 0 and 1 i.e. $0 \le K_c \le 1$.

Test of significance (for large samples i.e. n > 7):

If the number of observations for each of the variate is greater than 7, Chi-square approximation is used for testing the significance of coefficient of concordance in the population.

To test the hypothesis H_0: There exists no correlation between k variates based on ranks, the test statistic χ^2 is given by:

$$\chi^2 = k(n-1)K_c$$

Conclusion: If $\chi^2_{cal} \ge \chi^2_{tab}$ with (n-1) d.f. at chosen Ŭ, the null hypothesis is rejected. Otherwise, the null hypothesis is accepted.

Example-17: In a certain cattle judging competition, 10 cows were ranked by 4 judges (A, B, C and D) and ranks are given below. Compute the Kendallôs coefficient of concordance between the rankings of 4 judges and test for its significance.

	Cow									
Judge	1	2	3	4	5	6	7	8	9	10
A	5	3	4	1	8	7	6	9	10	2
B	4	6	3	2	7	5	10	8	9	1
C	4	7	5	3	6	9	8	10	2	1
D	3	5	1	4	9	10	7	6	8	2
Total (e_j)	16	21	13	10	30	31	31	33	29	6

Therefore, $\sum C_j = 220 \quad \sum C^2{}_j = 5754$

$$K_c = \frac{\sum_{j=1}^{n} C_j^{\,2} - \frac{(\sum C_j)^2}{n}}{\frac{1}{12} k^2 n\,(n^2 - 1)} \quad = \quad \frac{5754 - \frac{(220)^2}{10}}{\frac{1}{12}(4)^2 10\,(100 - 1)} = 0.69$$

Test of Significance:

To test the null hypothesis H_0 : There is no correlation between the ranking of 4 judges.

The value of test statistic is given by

$$\chi^2_{cal} = k\,(n-1)K_c = 4\ (10\text{-}\ 1)\ (0.69) = 24.84$$

Conclusion: Here χ^2_{cal} is more than critical value of χ^2 (= 16.92) with 9 d.f. at 5% level of significance. Hence, there exists significant correlation between the rankings of 4 judges.

EXERCISES

8g

1. A manufacturer of electric bulbs claims that he has developed a new production process which will increase the mean lifetime (00 hours) of bulbs from the present value 11.03. The results obtained from 15 bulbs taken at random from the new process are given below:

   11.29, 12.15, 10.69, 13.25, 13.47, 11.76, 14.05, 14.38, 11.08, 12.25, 10.93, 11.02, 12.87, 12.00, 13.56.

   Can it be concluded that mean life time of bulbs has increased by applying (i) Sign test (ii) Wilcoxon Signed Ranks Test.

2. Following is the arrangement of 25 men (m) and 15 women (w) lined up to purchase ticket for a picture show:

   m ww mmm w m w m w m w m ww

   mmm w mm ww mmmmm w m www mmmmmm

   Test randomness at 5% level of significance.

3. Nine animals were tested under control and experimental conditions and the following values were observed. Test whether there is significant increase in the values under experimental conditions by (i) Paired Sign Test (ii) Paired Sample Wilcoxon Signed Rank Test

Animal No.	1	2	3	4	5	6	7	8	9
Control	21	24	26	32	55	82	46	55	88
Experimental	18	27	35	42	82	99	52	30	62

4. Six students went on a diet in an attempt to loose weight with the following results

Student	A	B	C	D	E	F
Weight before (lbs)	174	191	188	182	201	188
Weight after (lbs)	165	186	183	178	203	181

8h

   Is the diet an effective means of loosing weight?

5. A sociologist asked twenty couples in the age group of 25-30 years regarding the number of children they planned to have. The answers of male and females were as follows:

Couple	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Male	3	2	1	2	3	4	3	0	2	3	5	6	4	2	1	2	3	5	6	3
Female	2	1	1	2	2	1	0	0	4	2	4	6	4	1	2	2	3	6	2	0

Apply sign test to conclude whether the attitude of male and female differ significantly on this issue.

6. A professor of sociology developed an intelligence test which he gave to two groups of 20 years old and 50 years old persons. The scores obtained were recorded in following table.

20 years old	**50 years old**
130	120
130	125
147	130
138	140
140	129
152	140
142	155
137	127
150	
140	

Perform the Mann-Whitney U-test to test the null hypothesis that average intelligence of both age groups is equal at 5% level.

7. Following are the test scores of 15 students of section I and 15 students of section II taught by two different professors. Test whether there is any significant difference in teaching method of two teachers by (i) Mann Whitney U-test (ii) Median test.

Roll No.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Scores of Section I	81	83	72	78	70	85	62	95	86	71	69	80	91	81	77
Scores of Section II	76	80	76	86	77	92	65	85	82	75	64	80	95	86	66

8. Yields from two varieties of wheat V_1 and V_2 sown on 6 and 7 identical plots are given below:

Variety	Yield (q/ha)						
V_1	40	35	52	60	46	55	
V_2	47	56	42	57	50	57	50

Use Mann-Whitney U test to test whether the two varieties have identical yields.

9. In order to compare the two varieties of maize, the following yields (kg) were recorded for ten identical plots under each variety:

Variety A	32.1	2.6	17.8	28.4	19.6	21.4	19.9	3.1	7.9	25.7
Variety B	19.8	27.6	23.7	9.9	3.8	27.6	34.1	18.7	16.9	17.9

i) Use Mann-Whitney U test to test whether the two varieties gave equal yields.

ii) Use Siegal-Tukey test to test whether the two maize varieties have the equal variability in yield.

10. To test the I.Q. of young and old people, an intelligence test was administered to two groups of 25 year old and 50 year old persons. The scores obtained by both groups of persons were obtained as given below:

	Scores									
25 years old	130	135	148	139	141	153	142	138	151	141
50 years old	122	126	131	142	130	141	156	128		

Using Mann-Whitney U-test, test whether the average performance of both age groups is same or not?