UNIT I4 TWO-SAMPLE TESTS

Structure

14.1 INTRODUCTION

In previous unit, we have already highlighted the main difference between parametric and non-parametric tests. Also we have discussed some well know/popular/commonly used one-sample non-parametric tests in the same unit. But comparison of a characteristic in two populations is a common problem faced by statisticians. We have already addressed this problem when distributions of the two populations are known to us. But what will do if assumption(s) of the test is (are) not fulfilled as distributions of the two populations are not known to us. This unit is devoted to address this problem. In this unit, we describe some of the frequently used two-sample tests based on non-parametric methods.

This unit is divided into seven sections. Section 14.1 is described the need of two-sample non-parametric tests. The paired sign test and Wilcoxon matched-pair signed-rank test for two samples are described in Sections 14.2 and 14.3 respectively. The Mann-Whitney U test which is used to test the difference in median of two independent populations is discussed in Section 14.4. In Section 14.5, we describe the Kolmogorov-Smirnov two-sample test. Unit ends by providing summary of what we have discussed in this unit in Section 14.6 and solution of exercises in Section 14.7.

Objectives

After completion of this unit, you should be able to:

- describe the need of two-sample non-parametric tests;
- identify which of the two-sample non-parametric tests is appropriate for a given situation;
- perform the test which is used in place of paired t-test when assumption(s) of the paired t-test is (are) not fulfilled, that is, paired sign test and Wilcoxon matched-pair signed-rank test;
- describe the Mann-Whitney U test which is used when the assumption(s) of two sample t-test is (are) not fulfilled;
- describe the Kolmogorov-Smirnor two-sample test for testing the hypothesis that two samples come from the populations having the same distribution.

14.2 PAIRED SIGN TEST

Paired sign test is the non-parametric version of the paired t-test. It is used when the assumption(s) of paired t-test is (are) not fulfilled.

In social sciences and in marketing research, generally two related groups are paired and we are interested to examine differences among two related groups. This often happens when the observations are recorded as before and after. Here, the paired observations are recorded on the same individuals or items. In these situations, we use paired t-test. If the assumptions of paired t-test is (are) not fulfilled then we use paired sign test, that means paired sign test is the non parametric version of the paired t-test. For example, if a college administrator is interested in knowing how a particular three-weeks training on basic Statistics affects student's grades in Statistics paper. In this situation, paired t-test is not applicable because of grades, that is, observations are given in ordinal scale as A, A^+, B, B^+, etc. But paired sign test can be applied because it needs only to convert the observations into plus and minus signs.

Assumptions

This test needs following assumptions to work:

(i) The pairs of measurements are independent.

(ii) The measurement scale is at least ordinal within each pair.

(iii) The variable under study is continuous.

Let us discuss the general procedure of this test:

Let $(X_1, Y_1), (X_2, Y_2), ..., (X_{n'}, Y_{n'})$ be a random sample of n' independent items / units / individuals, each with two measurements as before and after the training, diet, treatment, medicine, etc. Also let $\tilde{\mu}_1$ and $\tilde{\mu}_2$ be the medians of the populations before and after the training, diet, treatment, medicine, etc. Here, we want to test whether there is an effect of a diet, training, treatment, medicine, etc. or not. So we can take the null and alternative hypotheses as

$$H_0: \tilde{\mu}_1 = \tilde{\mu}_2 \text{ or } H_0: \tilde{\mu}_1 - \tilde{\mu}_2 = 0$$

$$H_1: \tilde{\mu}_1 \neq \tilde{\mu}_2 \text{ or } H_1: \tilde{\mu}_1 - \tilde{\mu}_2 \neq 0 \qquad [\text{for two-tailed test}]$$

or

$$\left.\begin{array}{l} H_0: \tilde{\mu}_1 \leq \tilde{\mu}_2 \text{ and } H_1: \tilde{\mu}_1 > \tilde{\mu}_2 \\ H_0: \tilde{\mu}_1 \geq \tilde{\mu}_2 \text{ and } H_1: \tilde{\mu}_1 < \tilde{\mu}_2 \end{array}\right\} \qquad [\text{for one-tailed test}]$$

The procedure for paired sign test is similar to sign test for one sample as discussed in pervious unit. After setting null and alternative hypothesis, paired sign test involves following steps:

Step 1: First of all, we convert the given observations into a sequence of plus and minus signs. For this, we find the difference $X_i - Y_i$ for all observations and consider only the sign of difference as plus or minus or compare X_i with Y_i for all observations. If $X_i > Y_i$ then we replace the pair of observation (X_i, Y_i) by a plus sign and if $X_i < Y_i$ then we replace (X_i, Y_i) by minus sign. But when $X_i = Y_i$, then corresponding pair of observations gives no information in terms of plus or minus signs so we exclude all such pairs from the analysis part. Due to such observations our common sample size reduces and let reduced common sample size be denoted by n.

Step 2: After that, we count the number of plus signs and number of minus signs. Suppose they are denoted by S^+ and S^- respectively.

Step 3: When null hypothesis is true and the population is dichotomised on the basis of sign of the difference then we expect that the number of plus signs (success) and number of minus signs (failure) approximately equal. And number of plus signs or number of minus signs follows binomial distribution (n, p = 0.5). For convenient consider the smaller number of plus or minus signs. If S^+ is less than S^- then we will take plus sign as success and minus sign as failure. Similarly, if S^- is less than S^+ then we will assume minus sign as success and plus sign as failure.

Step 4: To take the decision about the null hypothesis, we use concept of p-value as we have discussed in sign test for one sample. For p-value we determine the probability that test statistic is less than or equal to the actually observed of plus or minus signs. Since distribution of number of plus or minus signs is binomial (n, p = 0.5) therefore, this probability can be obtained with the help of **Table I** given in Appendix at the end of this block which provide the cumulative binomial probability and compare this probability with the level of significance (α). Here, test statistic depends upon the alternative hypothesis so the following cases arise:

The critical values of this test are not generally in tabular form and slightly difficult to obtain whereas p-value is easy to obtained with the help of cumulative binomial table so we use concept of p-value for take the decision about the null hypothesis.

For one-tailed test:

Case I: When $H_0: \tilde{\mu}_1 \le \tilde{\mu}_2$ and $H_1: \tilde{\mu}_1 > \tilde{\mu}_2$ (right-tailed test)

In this case, we expect that number of minus signs (S^-) is smaller than number of plus signs (S^+) therefore, the test statistic(S) is the number of minus signs (S^-). The p-value is determined as

$$\text{p-value} = P\left[S \le S^-\right]$$

If p-value is less than or equal to α, that is, $P\left[S \le S^-\right] \le \alpha$ then we reject the null hypothesis at α level of significance and if the p-value is greater than α then we do not reject the null hypothesis.

Case II: When $H_0: \tilde{\mu}_1 \ge \tilde{\mu}_2$ and $H_1: \tilde{\mu}_1 < \tilde{\mu}_2$ (left-tailed test)

In this case, we expect that number of plus signs (S^+) is smaller than number of minus signs (S^-) therefore, the test statistic(S) is the number of plus signs (S^+). The p-value is determined as

$$\text{p-value} = P\left[S \le S^+\right]$$

If p-value is less than or equal to α, that is, $P\left[S \le S^+\right] \le \alpha$ then we reject the null hypothesis at α level of significance and if the p-value is greater than α then we do not reject the null hypothesis.

For two-tailed test: When $H_0: \tilde{\mu}_1 = \tilde{\mu}_2$ and $H_1: \tilde{\mu}_1 \ne \tilde{\mu}_2$

For two tailed test, the test statistic (S) is the smaller of number of plus signs (S^+) and minus signs (S^-), that is,

$$S = \min\left\{S^+, S^-\right\}$$

and approximate p-value is determined as

$$\text{p-value} = 2P\left[S \le S^+\right]; \qquad \text{if } S^+ \text{ is small}$$

$$\text{p-value} = 2P\left[S \le S^-\right]; \qquad \text{if } S^- \text{ is small}$$

If p-value is less than or equal to α then we reject the null hypothesis at α level of significance and if the p-value is greater than α then we do not reject the null hypothesis.

For large sample $(n > 20)$:

For a large sample size n greater than and equal to 20, we use normal approximation to binomial distribution with mean

$$E(S) = np = n \times \frac{1}{2} = \frac{n}{2} \qquad \dots (1)$$

and variance

$$Var(S) = npq = n \times \frac{1}{2} \times \frac{1}{2} = \frac{n}{4} \qquad \dots (2)$$

Therefore, we use normal test (Z-test). The test statistic of Z-test is given by

$$Z = \frac{S - E(S)}{SE(S)}$$

$$= \frac{S - \frac{n}{2}}{\sqrt{\frac{n}{4}}} \sim N(0,1) \qquad \text{[Using equations (1) and (2)]} \qquad \dots (3)$$

After that, we calculate the value of test statistic Z and compare it with the critical value(s) given in **Table 10.1** at prefixed level of significance α. Take the decision about the null hypothesis as described in Section 10.2 of Unit 10 of this course.

Now, let us do some examples based on paired sign test.

Example 1: To compare the softness of the paper of two types A and B, a sample of each product of A and B is given to ten judges. Each judge rates the softness of each product on a scale from 1 to 10 with higher ratings implying a softer product. The results of the experiment are shown below. The first rating in the pair is of product A and the second is of product B.

(4, 5) (6, 4) (8, 5) (9, 8) (4, 1) (7, 9) (6, 2) (5, 3) (8, 2) (6, 7)

Apply the paired sign test to test the hypothesis that the products A and B are equally soft at 1% level of significance.

Solution: Here, the ratings are given in pair form so it can be considered as the case of dependent samples. But assumption of normality is not given also the given data are in the form of ranks so we cannot use the paired t-test in this case. So we go for paired sign test.

Here, we wish to test that the product A and B are equally soft. If $\tilde{\mu}_1$ and $\tilde{\mu}_2$ denote the average rank of softness of products A and B respectively then our claim is $\tilde{\mu}_1 = \tilde{\mu}_2$ and it complement is $\tilde{\mu}_1 \neq \tilde{\mu}_2$. Since claim contains the equality sign so we can take the claim as the null hypothesis and the complement as the alternative hypothesis. Thus,

$$H_0 : \tilde{\mu}_1 = \tilde{\mu}_2 \text{ and } H_1 : \tilde{\mu}_1 \neq \tilde{\mu}_2$$

Since the alternative hypothesis is two-tailed so the test is two-tailed test and the test statistic (S) is the minimum of plus signs (S^+) and minus signs (S^-), that is,

$$S = \min\left\{S^{+}, S^{-}\right\}$$

Calculation for S:

Judge	Product A	Product B	Sign of Difference (A−B)
1	4	5	–
2	6	4	+
3	8	5	+
4	9	8	+
5	4	1	+
6	7	9	–
7	6	2	+
8	5	3	+
9	8	2	+
10	6	7	–

From the above, we have

n = number of non zero differences = 10

S^{-} = number of plus signs = 7

S^{-} = number of minus signs = 3

Therefore,

$$S = \min\left\{S^{+}, S^{-}\right\} = \min\left\{7, 3\right\} = 3$$

To take the decision about the null hypothesis, we determine p-value with the help of **Table I** given in Appendix at the end of this block. Here, n = 10, p = 0.5 and r = 3. Thus, we have

$$\text{p-value} = 2P\left[S \le 3\right] = 2 \times 0.1719 = 0.3438$$

Since p-value = 0.3438 > 0.01(= α) so we do not reject the null hypothesis i.e. support the claim.

Thus, we conclude that the sample fails to provide us sufficient evidence against the claim so we may assume that products A and B are equally soft.

Example 2: The following are the weight (in pounds) before and after a dieting programme of 18 persons for one month:

Before	148.0	181.5	150.5	230.2	196.5	206.3	170.0	214.4	140.6
After	139.2	172.4	154.2	212.6	193.2	200.0	163.0	218.0	140.0
Before	166.8	150.3	197.2	168.0	172.4	160.3	210.0	172.4	180.2
After	166.8	156.2	185.4	158.5	167.2	162.3	202.3	166.2	170.5

Using the paired sign test to test whether the dieting programme is effective or not at 5% level of significance.

Solution: It is case of before and after study, so either we can go for paired t-test or paired sign test. But t-test requires assumptions of normality which is not given. So we go for paired sign test.

If $\tilde{\mu}_1$ and $\tilde{\mu}_2$ denote the median weights of the two depending populations before and after the dieting programme respectively. Here, we want to test whether the dieting programme is effective or not. So we can take our claim that the dieting programme is effective that means average (median) weight before dieting programme is greater than average (median) weight after the programme, that is, $\tilde{\mu}_1 > \tilde{\mu}_2$ and it complement is $\tilde{\mu}_1 \le \tilde{\mu}_2$. Since complement contains the equality sign so we can take the complement as the null hypothesis and the claim as the alternative hypothesis. Thus,

$H_0 : \tilde{\mu}_1 \le \tilde{\mu}_2$ and $H_1 : \tilde{\mu}_1 > \tilde{\mu}_2$

Since the alternative hypothesis is right-tailed so the test is right-tailed test.

Since alternative hypothesis is right-tailed so the test statistic (S) is the number of minus signs (S^-)

Calculation for S:

S. No	Weight before diet (X)	Weight after diet(Y)	Sign of Difference (X – Y)
1	148.0	139.2	+
2	181.5	172.4	+
3	150.5	154.2	–
4	230.2	212.6	+
5	196.5	193.2	+
6	206.3	200.0	+
7	170.0	163.0	+
8	214.4	218.0	–
9	140.6	140.0	+
10	166.8	166.8	Tie
11	150.3	156.2	–
12	197.2	185.4	+
13	168.0	158.5	+
14	172.4	167.2	+
15	160.3	162.3	–
16	210.0	202.3	+
17	172.4	166.2	+
18	180.2	170.5	+

From the above, we have

n = number of non-zero difference = 17

S^+ = number of plus signs = 13

S^- = number of minus signs = 4

Therefore, value of test statistic(S) is $S^- = 4$

To take the decision about the null hypothesis, we determine p-value with the help of **Table I** given in Appendix at the end of this block. Here, n = 17, p = 0.5 and r = 4. Thus, we have

$$\text{p-value} = P\left[S \le 4\right] = 0.0245$$

Since p-value = 0.0245 < 0.05 (α) so we reject the null hypothesis and support the alternative hypothesis i.e. we support the claim.

Thus, we conclude that the samples fail to provide us sufficient evidence against the claim so may assume that the dieting programme is effective.

Now, you can try the following exercises.

E1) Write one difference between paired sign test and paired t-test.

E2) Students are given a course designed to improve their IQ on a standard test. Their scores before and after the course are recorded as follows:

Before	95	112	128	96	97	117	105	95	99	86	75	82
After	99	115	133	96	99	120	106	106	100	89	90	85

With the help of paired sign test, examine whether IQ is affected due to course at 1% level of significance?

E3) A randomly chosen group of 15 consumers was asked to rate the two products A and B on a scale of 1 to 5 (5 being the highest). For each consumer, the first rating in the pair is of product A and the second is of product B. The data are:

(5, 2) (2, 3) (4, 1) (4, 2) (5, 3) (4, 4) (5, 4) (5, 1) (2, 4)
(3, 2) (3, 2) (4, 3) (3, 3) (3, 4) (2, 3)

The aim of study was to show that product A is favoured over product B. Do these data present evidence to justify this claim at 1% level of significance?

14.3 WILCOXON MATCHED-PAIR SIGNED-RANK TEST

We have already discussed Wilcoxon signed-rank test for one-sample in Unit 13. The Wilcoxon signed-rank test is also useful in comparing two populations for which we have paired observations. In this case, it is known as Wilcoxon matched-pair signed rank-test. As such, the test is a good alternative to the paired t-test in cases where the differences between paired observations are not believed to be normally distributed. Paired sign test discussed in previous section of this unit is also for the same situation. The paired sign test uses only the information whether X is larger or smaller than Y. This test is fine if the information about the observation sample is available in ordinal scale only. But if the measurement of the observation is available interval or ratio scales then choice of paired sign test is not recommended. Because this test do not take into account the information available in terms of the magnitude of the differences. To overcome this drawback of the paired sign test Wilcoxon matched-pair signed-rank test do the job for us, which takes into account the information of signs as well as of magnitude of differences. The Wilcoxon matched-pair signed-rank test is exactly the same as Wilcoxon signed-rank test described in Section 13.4 of previous unit of this course when $d_i = X_i - \tilde{\mu}_0$ is replaced by $d_i = X_i - Y_i$.

Assumptions

This test needs following assumptions to work:

(i) The distribution of differences d_i's is symmetric.

(ii) The differences d_i's are mutually independent.

(iii) The measurement scale is at least interval within each pair.

(iv) The variable under study is continuous.

Let us discuss the general procedure of this test:

Let $(X_1, Y_1), (X_2, Y_2), ..., (X_{n'}, Y_{n'})$ be a random sample of n' independent items / units / individuals each with two measurements as before and after the training, diet, treatment, medicine, etc. Also let $\tilde{\mu}_1$ and $\tilde{\mu}_2$ be the medians of the populations before and after the training, diet, treatment, medicine, etc. Here, we want to test whether there is an effect of a diet, training, treatment, medicine, etc. or not. So we can take the null and alternative hypotheses as

$$H_0: \tilde{\mu}_1 = \tilde{\mu}_2 \text{ or } H_0: \tilde{\mu}_1 - \tilde{\mu}_2 = 0$$

$$H_1: \tilde{\mu}_1 \neq \tilde{\mu}_2 \text{ or } H_1: \tilde{\mu}_1 - \tilde{\mu}_2 \neq 0 \quad [\text{for two-tailed test}]$$

or $\left.\begin{array}{l} H_0\text{: } \tilde{\mu}_1 \leq \tilde{\mu}_2 \text{ and } H_1\text{: } \tilde{\mu}_1 > \tilde{\mu}_2 \\ H_0\text{: } \tilde{\mu}_1 \geq \tilde{\mu}_2 \text{ and } H_1\text{: } \tilde{\mu}_1 < \tilde{\mu}_2 \end{array}\right\}$ [for one-tailed test]

The procedure of this test is similar to Wilcoxon signed-rank test for one-sample as discussed in pervious unit. After setting null and alternative hypotheses, this test involves following steps:

Step 1: First of all, we obtain the difference d_i for all observations with their plus and minus sign as

$$d_i = X_i - Y_i \text{ for all observations}$$

But when the observation X_i equal to Y_i then the difference $d_i = X_i - Y_i$ gives no information in terms of signs as well as in terms of magnitude of the difference. So we exclude all such observations from the analysis part. Due to these observations our sample size reduces and let reduced sample size be denoted by n.

Step 2: After that, we find the absolute value of these d_i's obtained in Step1as $|d_1|, |d_2|, \ldots, |d_n|$.

For more detailed about repeated (tied) rank please go through Section 7.4 of the Unit 2 of MST-002

Step 3: In this step, we rank all $|d_i|$'s (obtained in Step 2) with respect to their magnitudes from smallest to largest, that is, the rank 1 is given to the smallest of $|d_i|$'s, rank 2 is given to the second smallest and so on up to the largest $|d_i|$'s. If several values are same (tied), we assign each the average of ranks they would have received if there were no repetition.

Step 4: Now, assign the signs to the ranks which the original differences have.

Step 5: Finally, we calculate the sum of the positive ranks (T^+) and sum of negative ranks (T^-) separately.

Under H_0, we expect approximately equal number of positive and negative ranks. And under the assumption that distribution of differences is symmetric about its median, we expect that sum of the positive ranks (T^+) and sum of negative ranks (T^-) are equal.

Step 6: **Decision Rule:**

To take the decision about the null hypothesis, the test statistic (T) is the smaller of T^+ and T^-. And the test statistic is compared with the critical (tabulated) value for a given level of significance (α) under the condition that the null hypothesis is true. **Table II** given in Appendix at the end of this block provides the critical values of test statistic at α level of significance for both one-tailed and two-tailed tests. Here, test statistic depends upon the alternative hypothesis so the following cases arise:

For one-tailed test:

Case I: When $H_0 : \tilde{\mu}_1 \leq \tilde{\mu}_2$ and $H_1 : \tilde{\mu}_1 > \tilde{\mu}_2$ (right-tailed test)

In this case, we expect that sum of negative ranks (T^-) is smaller than sum of positive ranks (T^+) therefore, the test statistic (T) is the sum of negative ranks (T^-).

If computed value of test statistic (T) is less than or equal to the critical value T_α at α level of significance, that is, $T \le T_\alpha$ then we reject the null hypothesis at α level of significance, otherwise we do not reject the null hypothesis.

Case II: When $H_0 : \tilde{\mu}_1 \ge \tilde{\mu}_2$ and $H_1 : \tilde{\mu}_1 < \tilde{\mu}_2$ (left-tailed test)

In this case, we expect that sum of positive ranks (T^+) is smaller than sum of negative ranks (T^-) therefore, the test statistic (T) is the sum of positive ranks (T^+).

If computed value of test statistic (T) is less than or equal to the critical value T_α at α level of significance, that is, $T \le T_\alpha$ then we reject the null hypothesis at α level of significance, otherwise we do not reject the null hypothesis.

For two-tailed test:

When $H_0 : \tilde{\mu}_1 = \tilde{\mu}_2$ and $H_1 : \tilde{\mu}_1 \ne \tilde{\mu}_2$

In this case, the test statistic (T) is the smaller of sum of positive ranks (T^+) and sum of negative ranks (T^-), that is,

$$T = \min\left\{T^+, T^-\right\}$$

If computed value of test statistic (T) is less than or equal to the critical value $T_{\alpha/2}$ at α level of significance, that is, $T \le T_{\alpha/2}$ then we reject the null hypothesis at α level of significance, otherwise we do not reject the null hypothesis.

For large sample (n > 25):

For a large sample size n greater than 25, the distribution of test statistic (T) approximated by a normal distribution with mean

$$E(T) = \frac{n(n+1)}{4} \qquad \ldots (4)$$

and variance

$$\mathrm{Var}(T) = \frac{n(n+1)(2n+1)}{24} \qquad \ldots (5)$$

The proof of mean and variance of test statistic T is beyond the scope of this course.

Therefore in this case, we use normal test (Z-test). The test statistic of Z-test is given by

$$Z = \frac{T - E(T)}{SE(T)} = \frac{T - E(T)}{\sqrt{\mathrm{Var}(T)}}$$

$$= \frac{T - \dfrac{n(n+1)}{4}}{\sqrt{\dfrac{n(n+1)(2n+1)}{24}}} \sim N(0,1) \left[\begin{array}{c}\text{Using equations}\\ \text{(4) and (5)}\end{array}\right] \qquad \ldots (6)$$

After that, we calculate the value of test statistic Z and compare it with the critical value(s) given in **Table 10.1** at prefixed level of significance α. Take the decision about the null hypothesis as described in Section 10.2 of Unit 10 of this course.

Now, it is time to do some examples based on above test.

Example 3: A group of 12 children was tested to find out how many digits they would repeat from memory after hearing them once. They were given practice session for this test. Next week they were retested. The results obtained were as follows:

Child Number	1	2	3	4	5	6	7	8	9	10	11	12
Recall Before	6	4	5	7	6	4	3	7	8	4	6	5
Recall After	6	6	4	7	6	5	5	9	9	7	8	7

Assuming that the distribution of the differences of the scores before and after the practice session is symmetrical about its median, can the memory practice session improve the performance of children at 5% level of significance using Wilcoxon matched-pair signed-rank?

Solution: It is case of before and after study so we can go for paired t-test or paired sign test or Wilcoxon matched-pair signed-rank. But t-test requires assumptions of normality which is not given. Also measurement of observations is available in interval scale, so we will go for Wilcoxon matched-pair signed-rank test.

Here, we wish to test that memory practice session improves the performance of children. If $\tilde{\mu}_1$ and $\tilde{\mu}_2$ denote the average (median) score before and after the practice session so our claim is $\tilde{\mu}_1 < \tilde{\mu}_2$ and its complement is $\tilde{\mu}_1 \geq \tilde{\mu}_2$. Since complement contains the equality sign so we can take the complement as the null hypothesis and the claim as the alternative hypothesis. Thus,

$$H_0 : \tilde{\mu}_1 \geq \tilde{\mu}_2$$

$$H_1 : \tilde{\mu}_1 < \tilde{\mu}_2 \qquad \left[\text{memory practice session improves the performance}\right]$$

Since the alternative hypothesis is left-tailed so the test is left-tailed test and the test statistic (T) is the sum of positive ranks (T^+).

Calculation for T:

Child Number	**Recall Before (X)**	**Recall After (Y)**	**Difference (X−Y)**	**Absolute Difference $\|d\|$**	**Rank of $\|d\|$**	**Signed Rank**
1	6	6	Tie	---	----	----
2	4	6	−2	2	6	− 6
3	5	4	−1	1	2	−2
4	7	7	Tie	---	----	----
5	6	6	Tie	---	----	----
6	4	5	1	1	2	2
7	3	5	−2	2	6	− 6
8	7	9	−2	2	6	− 6
9	8	9	−1	1	2	−2
10	4	7	−3	3	9	− 9
11	6	8	−2	2	6	− 6
12	5	7	−2	2	6	− 6

From the above calculations, we have

T^+ = sum of positive ranks = 2

n = number of non-zero d_i's = 9

So the value of test statistic (T) is $T^+ = 2$

The critical (tabulated) value of test statistic for one-tailed test corresponding n = 9 at 5% level of significance is 9.

Since calculated value of test statistic T(= 2) is less than the critical value (= 9) so we reject the null hypothesis and support the alternative hypothesis i.e. we support the claim at 5% level of significance.

Thus, we conclude that samples fail to provide us sufficient evidence against the claim so we may assume that memory practice session improves the performance of children.

Example 4: A drug is given to 30 patients and the increment in their blood pressures were recorded to be 2, 0, − 4, −1, 2, 3, −2, 5, − 6, 6, 2, 4, 3, 5, 4, 2, 2, −1, −3, −5, 4, 6, −8, − 4, 2, 4, 0, − 6, 3, 1. Assuming that the distribution of the increment in the blood pressures is symmetrical about its median, is it reasonable to believe that the drug is effective to reduce the blood pressure at 1% level of significance?

Solution: We are given that the increment in blood pressures of 30 patients after given a drug to the patients. The assumption of normality for increment in blood pressure is not given so we will not go for paired t-test. Also measurement of observations is available in interval scale, so we will go for Wilcoxon matched pair signed-rank test.

If $\tilde{\mu}_1$ and $\tilde{\mu}_2$ denote medians of blood pressures of two populations before and after the drug respectively and we want to test that drug is effective to reduce the blood pressure so our claim is $\tilde{\mu}_1 > \tilde{\mu}_2$ and its complement is $\tilde{\mu}_1 \le \tilde{\mu}_2$. Since complement contains the equality sign so we can take the complement as the null hypothesis and the claim as the alternative hypothesis. Thus,

$$H_0 : \tilde{\mu}_1 \le \tilde{\mu}_2$$

$$H_1 : \tilde{\mu}_1 > \tilde{\mu}_2 \qquad \left[\text{The drug is efective}\right]$$

Since the alternative hypothesis is right-tailed so the test is right-tailed test.

Since in this exercise the increment in the blood pressures is given that means we are given that the difference of after on before that is **Y − X** instead of **X – Y** therefore, for testing the null hypothesis, the test statistic is the sum of positive ranks (T^+) instead of sum of negative ranks (T^-).

Calculation for T:

Patient Number	**Increment $(d = X - Y)$**	**Absolute Difference $\lvert d \rvert$**	**Rank of $\lvert d \rvert$**	**Signed Rank**
1	2	2	7	7
2	0	Tie	---	---
3	−4	4	17.5	−17.5
4	−1	1	2	−2
5	2	2	7	7
6	3	3	12.5	12.5
7	−2	2	7	−7
8	5	5	22	22
9	−6	6	25.5	−25.5
10	6	6	25.5	25.5
11	2	2	7	7
12	4	4	17.5	17.5
13	3	3	12.5	12.5

14	5	5	22	22
15	4	4	17.5	17.5
16	2	2	7	7
17	2	2	7	7
18	−1	1	2	−2
19	−3	3	12.5	−12.5
20	−5	5	22	−22
21	4	4	17.5	17.5
22	6	6	25.5	25.5
23	−8	8	28	−28
24	4	4	17.5	17.5
25	2	2	7	7
26	4	4	17.5	17.5
27	0	Tie	---	---
28	−6	6	25.5	−25.5
29	3	3	12.5	12.5
30	1	1	2	2

From the above calculations, we have

$T^+ = 264$

n = number of non-zero d_i's = 28

Therefore, the value of test statistic (T) is $T^+ = 264$

Also, n = 28 (> 25) therefore, it is the case of large sample. So in this case, we use Z-test. The test statistic of Z-test is given by

$$Z = \frac{T - E(T)}{SE(T)} \sim N(0,1)$$

where, $E(T) = \frac{n(n+1)}{4} = \frac{28(28+1)}{4} = 203$ and

$$SE(T) = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{28(28+1)(2\times 28+1)}{24}} = 43.91$$

Putting the values in test statistic Z, we have

$$Z = \frac{264 - 203}{43.91} = 1.38$$

The critical (tabulated) value of test statistic for right-tailed test at 1% level of significance is $z_\alpha = z_{0.01} = 1.645$.

Since calculated value of Z (= 1.38) is less than the critical value (= 1.645), that means it lies in non-rejection region, so we do not reject the null hypothesis and reject the alternative hypothesis i.e. we reject the claim at 1% level of significance.

Thus, we conclude that the sample provide us sufficient evidence against the claim so we can say that drug is not effective to reduce the blood pressure.

Now, you can try the following exercises.

E4) Write one assumption of the Wilcoxon matched-pair singed-rank test which is not required in paired sign test.

E5) A group of 10 students were given a test in Statistics and after one month's coaching they were given another test of the similar nature. The following table gives the increment in their marks in the second test over the first:

Roll No.	1	2	3	4	5	6	7	8	9	10
Increase in Marks	6	−2	8	−4	10	2	5	−4	6	0

Assuming that the distribution of the increment in the marks is symmetrical about its median, do the marks indicate that students have benefited by coaching. Test it at 1% level of significance?

E6) To compare the softness of the paper products, a sample of two products A and B is given to ten judges. Each judge rates the softness of each product on a scale from 1 to 10 with higher ratings implying a softer product. The results of the experiment are shown below. The first rating in the pair is of product A and the second is of product B.

(4, 5) (6, 4) (8, 5) (9, 8) (4, 1) (7, 9) (6, 2) (5, 3) (8, 2) (6, 7)

Assume that the distribution of the differences of the ranks of products A and B is symmetrical about its median. Apply the Wilcoxon matched pair signed-rank test to test the hypothesis that the products A and B are equally soft at 1% level of significance.

14.4 Mann-Whitney U Test

In Section 14.2 and 14.3, we have discussed the paired sign test and Wilcoxon matched-pair signed-rank test respectively which are used when the observations are dependent, that is, observations are paired. When we are interested in testing of difference in means of two independent populations then we use two sample t-test. To use the t-test however, it is necessary to make a set of assumptions (as described in Unit 11 of this course). In particular, it is necessary that two independent samples be randomly drawn from normal populations having equal variances and the data be measured in at least of an interval scale. But in studies of consumer behaviour, marketing research, experiment of psychology, etc. generally the data are collected in ordinal scale and the form of the population is not known. Since the parametric t-test could not be used in such situation, an appropriate non-parametric technique is needed. In such a circumstance a very simple non-parametric test known as Mann-Whitney U test may be used.

This test was developed jointly by Mann, Whitney and Wilcoxon. Wilcoxon considered only the case of equal sample sizes while Mann and Whitney seem to have been the first to treat the case of unequal sample sizes. Therefore, it is sometimes called the Mann-Whitney U test and sometimes the Wilcoxon ranks sum test. Thus, this test may also be viewed as a substitute for the parametric t-test for the difference between two population means. When assumptions of the two-sample t-test are fulfilled then this test is slightly weaker than t-test.

Assumptions

This test work under the following assumptions:

(i) The two samples are randomly and independently drawn from their respective populations.

(ii) The variable under study is continuous.

(iii) The measurement scale is at least ordinal.

(iv) The distributions of two populations differ only with respect to location parameter.

Let us discuss the general procedure of this test:

Let us suppose that we have two independent random samples $X_1, X_2, ..., X_{n_1}$ and $Y_1, Y_2, ..., Y_{n_2}$ drawn from two populations having medians $\tilde{\mu}_1$ and $\tilde{\mu}_2$ respectively. Here, we want to test the hypothesis about the medians of two populations so we can take the null and alternative hypotheses as

$$H_0: \tilde{\mu}_1 = \tilde{\mu}_2 \text{ and } H_1: \tilde{\mu}_1 \neq \tilde{\mu}_2 \qquad [\text{for two-tailed test}]$$

or

$$\left.\begin{array}{l} H_0: \tilde{\mu}_1 \leq \tilde{\mu}_2 \text{ and } H_1: \tilde{\mu}_1 > \tilde{\mu}_2 \\ H_0: \tilde{\mu}_1 \geq \tilde{\mu}_2 \text{ and } H_1: \tilde{\mu}_1 < \tilde{\mu}_2 \end{array}\right\} \qquad [\text{for one-tailed test}]$$

We can also form the null and alternative hypotheses in the form of distribution functions. If $F_1(x)$ and $F_2(x)$ are the distribution functions of the first and second populations respectively and we want to test that the two independent samples come from the populations that are identical with respect to location, that is, median.

Therefore, we can take the null and alternative hypotheses as

$$\left.\begin{array}{ll} H_1: F_1(x) = F_2(x) & \text{for at least one value of } x \\ H_1: F_1(x) \neq F_2(X) & \text{for at least one value of } x \end{array}\right\} \qquad [\text{for two-tailed test}]$$

or

$$\left.\begin{array}{ll} H_0: F_1(x) \leq F_2(x) & \text{for at least one value of } x \\ H_1: F_1(x) > F_2(X) & \text{for at least one value of } x \end{array}\right\} \qquad [\text{for right-tailed test}]$$

or

$$\left.\begin{array}{ll} H_0: F_1(x) \geq F_2(x) & \text{for at least one value of } x \\ H_1: F_1(x) < F_2(X) & \text{for at least one value of } x \end{array}\right\} \qquad [\text{for left-tailed test}]$$

After setting null and alternative hypotheses, this test involves following steps:

Step 1: First of all, we combine the observations of two samples.

Step 2: After that, ranking all these combined observations from smallest to largest, that is, the rank 1 is given to the smallest of the combined observations, rank 2 is given to the second smallest and so on up to the largest observation. If several values are same (tied), we assign each the average of ranks they would have received if there were no repetition.

Step 3: If null hypothesis is true then we can expect that the sum of ranks of the two samples are equal. Now for convenience, we consider the smaller sized sample and calculate the sum of ranks of the observations of this sample. Let S be the sum of the ranks assigned to the sample observations of smaller sized sample then for testing the null hypothesis the test statistic is given by

$$U = S - \frac{n_1(n_1+1)}{2}; \quad \text{if } n_1 \text{ is small}$$

$$U = S - \frac{n_2(n_2+1)}{2}; \quad \text{if } n_2 \text{ is small}$$

Step 4: Obtain critical value(s) of test statistic U at given level of significance under the condition that null hypothesis is true. **Table VII** in the Appendix at the end of this block provides the lower and upper critical values for a given combination of n_1 and n_2 at α level of significance for one-tailed and two-tailed test.

Step 5: Decision rule:

To take the decision about the null hypothesis, the calculated value of test statistic U (computed in Step 3) is compared with the critical (tabulated) value (obtained in Step 4) at a given level of significance (α) under the condition that null hypothesis is true. Since test may be one-tailed or two-tailed so following cases arise:

For two-tailed test: When $H_0 : \tilde{\mu}_1 = \tilde{\mu}_2$ and $H_1 : \tilde{\mu}_1 \neq \tilde{\mu}_2$

For two-tailed test, we see critical values at α/2 for α level of significance. If calculated value of test statistic U is either less than or equal to the lower critical value $(U_{L,\alpha/2})$ or greater than or equal to the upper critical value $(U_{U,\alpha/2})$, that is, $U \le U_{L,\alpha/2}$ or $U \ge U_{U,\alpha/2}$ then we reject the null hypothesis at α level of significance. However, if computed U lies between these critical values, that is, $U_{L,\alpha/2} < U < U_{U,\alpha/2}$, then we do not reject the null hypothesis at α% level of significance.

For one-tailed test:

For one-tailed test, we see the critical value at α for α level of significance.

Case I: When $H_0 : \tilde{\mu}_1 \le \tilde{\mu}_2$ and $H_1 : \tilde{\mu}_1 > \tilde{\mu}_2$ (right-tailed test)

If calculated value of test statistic U is greater than or equal to the upper critical value $(U_{U,\alpha})$, that is, $U \ge U_{U,\alpha}$ then we reject the null hypothesis at α level of significance. However, if computed U is less than upper critical value $(U_{U,\alpha})$, that is, $U < U_{U,\alpha}$ then we do not reject the null hypothesis at α% level of significance.

Case II: When $H_0 : \tilde{\mu}_1 \ge \tilde{\mu}_2$ and $H_1 : \tilde{\mu}_1 < \tilde{\mu}_2$ (left-tailed test)

If calculated value of test statistic U is less than or equal to the lower critical value $(U_{L,\alpha})$, that is, $U \le U_{L,\alpha}$ then we reject the null hypothesis at α level of significance. However, if computed U is greater than lower critical value $(U_{L,\alpha})$, that is, $U > U_{L,\alpha}$ then we do not reject the null hypothesis at α% level of significance.

For large (n_1 or n_2 >20):

When either n_1 or n_2 exceeds 20, the statistic U is approximately normally distributed with mean

$$E(U) = \frac{n_1 n_2}{2} \qquad \dots (7)$$

and variance

$$\text{Var}(U) = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12} \qquad \dots (8)$$

The proof of mean and variance of test statistic(T) is beyond the scope of this course.

Therefore in this case, we use normal test (Z-test) (described in Unit 10 of Block 3 of this course.). The test statistic of Z-test is given by

$$Z = \frac{U - E(U)}{SE(U)} = \frac{U - E(U)}{\sqrt{\text{Var}(U)}} \sim N(0, 1)$$

$$= \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} \sim N(0,1) \left[\begin{array}{c}\text{Using equations} \\ \text{(7) and (8)}\end{array}\right] \qquad \ldots (9)$$

After that, we calculate the value of test statistic Z and compare it with the critical value(s) given in **Table 10.1** at prefixed level of significance α. Take the decision about the null hypothesis as described in Section 10.2 of Unit 10 of this course.

Let us do some examples to become more user friendly with the test explained above:

Example 5: A Statistics professor taught two special sections of a basic course in which students in each section were considered outstanding. He used a "traditional" method of instruction (T) in one section and an "experimental" method of instruction (E) in the other. At the end of the semester, he ranked the students based on their performance from 1 (worst) to 20 (best).

T	1	2	3	5	8	10	12	13	14	15
E	4	6	7	9	11	16	17	18	19	20

Test whether there is any evidence of a difference in performances based on the two methods at 5% level of significance.

Solution: It is case of two independent populations and the assumption of normality of both the populations is not given also the given data are in the form of ranks so we cannot use two-sample t-test in this case. Also sample sizes are small so we cannot use the Z-test. Therefore, we go for Mann-Whitney U test.

Here, we want to test that performance of the students based on the two methods is different. If $\tilde{\mu}_1$ and $\tilde{\mu}_2$ denote the average (median) ranks of the students who taught by traditional" and "experimental" methods respectively so our claim is $\tilde{\mu}_1 \neq \tilde{\mu}_2$ and its complement is $\tilde{\mu}_1 = \tilde{\mu}_2$. Since complement contains the equality sign so we can take the complement as the null hypothesis and the claim as the alternative hypothesis. Thus,

$$H_0 : \tilde{\mu}_1 = \tilde{\mu}_2$$

$$H_1 : \tilde{\mu}_1 \neq \tilde{\mu}_2 \qquad [\text{both methods are different}]$$

Since the alternative hypothesis is two-tailed so the test is two-tailed test.

Here, combined ranks are given. Also $n_1 = n_2$ so we take any sum of ranks(S) of first sample or second sample. We consider first sample. Thus,

$$S = 1 + 2 + 3 + 5 + 8 + 10 + 12 + 13 + 14 + 15 = 83$$

Here, we consider the first sample so the test statistic is given by

$$U = S - \frac{n_1(n_1+1)}{2} = 83 - \frac{10 \times (10+1)}{2} = 28$$

Since test is two-tailed so the lower and upper critical values of test statistic for two-tailed test corresponding $n_1 = 10$ and $n_2 = 10$ at 5% level of significance are $U_{L,\alpha/2} = U_{L,\,0.025} = 79$ and $U_{U,\alpha/2} = U_{U,\,0.025} = 131$.

Since calculated value of test statistic U (= 28) is less than critical values (= 79 and 131) so we reject the null hypothesis and support the alternative hypothesis i.e. we support the claim at 5% level of significance.

Thus, we conclude that the samples fail to provide us sufficient evidence against the claim so we can say that performances based on two methods are different.

Example 6: The following data show the median amount to be spent on a birthday gift by students who are unemployed and employed:

Unemployed	200	350	500	250	150	275			
Employed	550	500	1000	450	700	200	500	600	1000

Examine that the median amount to be spent on a birthday gift by students who are unemployed is lower than for students who are employed by Mann-Whitney U test at 1% level of significance.

Solution: It is case of two independent populations and the assumption of normality of both the populations is not given so we cannot use t-test in this case. Also sample sizes are small so we cannot use the Z-test. Therefore, we go for Mann-Whitney U test.

Here, we want to test that the median amount to be spent on a birthday gift by students who are unemployed is lower than for students who are employed. If $\tilde{\mu}_1$ and $\tilde{\mu}_2$ denote the average (median) amount to be spent on a birthday gift by students who are unemployed and employed respectively so our claim is $\tilde{\mu}_1 < \tilde{\mu}_2$ and its complement is $\tilde{\mu}_1 \geq \tilde{\mu}_2$. Since complement contains the equality sign so we can take the complement as the null hypothesis and the claim as the alternative hypothesis. Thus,

$$H_0 : \tilde{\mu}_1 \geq \tilde{\mu}_2$$

$$H_1 : \tilde{\mu}_1 < \tilde{\mu}_2 \begin{bmatrix} \text{median amount to be spent by students who are} \\ \text{unemployed is lower than the students who are employed} \end{bmatrix}$$

Since the alternative hypothesis is left-tailed so the test is left-tailed test.

To perform the Mann-Whitney U test, we combine the observations of two samples and for convenient, we arrange the data in ascending order. After that we rank all these combined observations from smallest to largest as:

Unemployed	**Ranks**	**Unemployed**	**Ranks**
150	1	200	2.5
200	2.5	450	7
250	4	500	9
275	5	500	9
350	6	550	11
500	9	600	12
		700	13
		1000	14.5
		1000	14.5
Total	27.5		

Since n_1 (= 6) is small than n_2 (= 9) so we take sum of the ranks(S) assigned to the sample of unemployed students. Thus, S = 27.5.

Here, we consider the first sample so the test statistic is given by

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

$$= 27.5 - \frac{6\times(6+1)}{2} = 6.5$$

To decide about the null hypothesis, the calculated value of test statistic U is compared with the lower critical (tabulated) value at 1% level of significance.

Since test is left-tailed so the lower critical value of test statistic for left-tailed corresponding to $n_1 = 6$ and $n_2 = 9$ at 1% level of significance is $U_{L,\alpha} = U_{L,\,0.01} = 29$.

Since calculated value of test statistic U (= 6.5) is less than lower critical value (= 29) so we reject the null hypothesis and support the alternative hypothesis i.e. we support the claim at 1% level of significance.

Thus, we conclude that the samples fail to provide us sufficient evidence against the claim so we may assume that the median amount to be spent on a birthday gift by students who are unemployed is lower than for students who are employed.

Now, you can do following exercises in same manner.

E7) Write one difference between two-sample t-test and Mann-Whitney U test.

E8) Write one difference between Wilcoxon matched-pair singed-rank test and Mann-Whitney U test.

E9) The senior class in a particular high school had 25 boys. Twelve boys lived in villages and other thirteen lived in a town. A test was conduct to see that village boys in general were physically fit than the town boys. Each boy in the class was given a physical fitness test in which a low score indicates poor physical condition. The scores of the village boys (V) and the town boys(T) are as follows:

Village Boys(V)	15.7	8.2	6.5	7.2	9.0	4.5	10.6
	12.4	16.2	12.9	11.4	5.6		
Town Boys(T)	12.7	3.2	11.8	7.9	5.6	6.7	12.6
	7.9	2.7	6.1	3.6	6.5	2.8	

Test whether the village boys are more fit than town boys at 5% level of significance.

E10) The following data represent lifetime (in hours) of batteries for two different brands A and B:

Brand A	40	30	55	40	40	35	30	40	50	45	40	35
Brand B	45	60	50	60	35	50	55	60	50	50	40	55

Examine, is the average life of two brands is same at 1% level of significance?

14.5 KOLMOGOROV-SMIRNOV TWO-SAMPLE TEST

In Mann-Whitney U test, we have tested the hypothesis that the two independent samples come from the populations that are identical with respect to location whereas Kolmogorov-Smirnov two-sample test is sensitive to differences of all types that may exist between two distributions, that is, location, dispersion, skewness, etc. Therefore, it is referred as a general test. This test was developed by Smirnov. This test also carries the name of Kolmogorov because of its similarity to the one-sample test developed by Kolmogorov. In Kolmogorov-Smirnov one sample test, the observed (sample or empirical) cumulative distribution function is compared with the hypothesized cumulative distribution function whereas in two-sample case the comparison is made between the empirical cumulative distributions functions of the two samples.

Assumptions

The assumptions necessary for this test are:

(i) The two samples are randomly and independently drawn from their respective populations.

(ii) The variable under study is continuous.

(iii) The measurement scale is at least ordinal.

Let us discuss the general procedure of this test:

Let us suppose that we have two independent random samples $X_1, X_2, ..., X_{n_1}$ and $Y_1, Y_2, ..., Y_{n_2}$ drawn from first and second populations with distribution functions $F_1(x)$ and $F_2(x)$ respectively. Also let $S_1(x)$ and $S_2(x)$ be the sample or empirical cumulative distribution functions of samples drawn from first and second populations respectively.

Generally, we want to test whether independent random samples come from populations having the same distribution functions in all respect or the distribution functions of two populations differ with respect to location, dispersion, skewness, etc. So we can take the null and alternative hypotheses as

$$H_0 : F_1(x) = F_2(x) \quad \text{for all values of x}$$

$$H_1 : F_1(x) \neq F_2(x) \quad \text{for at least one value of x}$$

After setting null and alternative hypotheses, the procedure of Kolmogorov - Smirnov two-sample test summarise in the following steps:

Step 1: This test is based on the comparison of the empirical (sample) cumulative distribution functions, therefore, first of all we compute sample cumulative distribution functions $S_1(x)$ and $S_2(x)$ from the sample data as the proportion of the number of sample observations less than or equal to some number x to the number of observations, that is,

$$S_1(x) = \frac{\text{The number of sample observations less than or equal to x}}{n_1}$$

and

$$S_2(x) = \frac{\text{The number of sample observations less than or equal to x}}{n_2}$$

Step 2: After finding the empirical cumulative distribution functions $S_1(x)$ and $S_2(x)$ for all possible values of x, we find the deviation between the empirical cumulative distribution functions for all x. That is,

$$S_1(x) - S_2(x) \quad \text{for all x}$$

Step 3: If the two samples have been drawn from identical populations then $S_1(x)$ and $S_2(x)$ should be fairly close for all value of x. Therefore, we find the point at which the two functions show the maximum deviation. So we take the test statistic which calculate the greatest vertical deviation between $S_1(x)$ and $S_2(x)$, that is,

$$D = \sup_x |S_1(x) - S_2(x)|$$

which is read as "D equals the supreme over all x, of the absolute value of the difference $S_1(x) - S_2(x)$"

Step 4: Obtain critical value of test statistic at α % level of significance under the condition that null hypothesis is true. **Table V and VI** in Appendix provides the critical values of test statistic for equal and unequal sample sizes respectively at different level of significance.

Step 5: **Decision Rule:**

To take the decision about the null hypothesis, the test statistic (calculated in Step 3) is compared with the critical (tabulated) value (obtained in Step 4) for a given level of significance (α).

If computed value of D is greater than or equal to critical value (D_α), that is, $D \geq D_\alpha$ then we reject the null hypothesis at α level of significance, otherwise we do not reject H_0.

For large sample ($n_1 = n_2 = n > 40$ and $n_1 > 16$ for unequal):

For a sample size $n_1 = n_2 = n > 40$ and $n_1 > 16$ for unequal sample sizes, the critical value of test statistic at given level of significance is approximated by the formula given in last row of **Table V & VI** in the Appendix.

Now, it is time to do some examples based on above test.

Example 7: The following data represent lifetime (in hours) of batteries for two different brands A and B:

Brand A	40	30	55	40	40	35	30	40	50	45	40	35
Brand B	45	60	50	60	35	50	55	60	50	50	40	55

Examine by the Kolmogorov-Smirnov two-sample test, is the average life of two brands same at 5% level of significance?

Solution: Here, we want to test that the average life of batteries of two brands is same. If $F_1(x)$ and $F_2(x)$ are cumulative distribution functions of the life of batteries of brand A and brand B respectively then our claim is $F_1(x) = F_2(x)$ and its complement is $F_1(x) \neq F_2(x)$. Since the claim contains the equality sign so we can take claim as the null hypothesis and complement as the alternative hypothesis. Thus,

$$H_0 : F_1(x) = F_2(x) \quad \text{for all values of x}$$

$$H_1 : F_1(x) \neq F_2(x) \quad \text{for at least one value of x}$$

So the test statistic is given by

$$D = \sup_x |S_1(x) - S_2(x)|$$

where, $S_1(x)$ and $S_2(x)$ are the empirical cumulative distribution functions of the lives of the batteries of brand A and brand B drawn in the samples respectively which are defined as:

$$S_1(x) = \frac{\text{number of batteries of brand A whose average life} \leq x}{n_1}$$

$$S_2(x) = \frac{\text{number of batteries of brand B whose average life} \leq x}{n_2}$$

Calculation for $|S_1(x) - S_2(x)|$:

Average Life	**Frequency (Brand A)**	**C.F. Brand A**	**$S_A(x)$**	**Frequency (Brand B)**	**C.F. Brand B**	**$S_B(x)$**	$\lvert S_1(x) - S_2(x)\rvert$
30	2	2	2/12	0	0	0	2/12
35	2	4	4/12	1	1	1/12	3/12
40	5	9	9/12	1	2	2/12	7/12
45	1	10	10/12	1	3	3/12	7/12
50	1	11	11/12	4	7	7/12	4/12
55	1	12	12/12	2	9	9/12	3/12
60	0	12	12/12	3	12	12/12	0
Total	$n_1 = 12$			$n_2 = 12$			

From the above calculation, we have

$$D = \sup_x |S_1(x) - S_2(x)| = \frac{7}{12} = 0.58$$

The critical value of test statistic for equal sample sizes $n_1 = n_2 = 12$ at 5% level of significance is 6/12 = 0.5.

Since calculated value of test statistic (= 0.58) is greater than critical value (= 0.5) so we reject the null hypothesis i.e. we reject the claim at 5% level of significance.

Thus, we conclude that the samples provide us sufficient evidence against the claim so the average life of batteries of two brands is different.

Example 8: The following are the marks in Statistics of B.Sc. students taken randomly from two colleges A and B:

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
College (A)	2	2	4	6	3	3	4	8	7	5
College (B)	1	1	2	5	7	3	3	2	6	6

Apply Kolmogorov-Smirnov test to examine that the distribution of marks in college A and college B is same at 1% level of significance.

Solution: Here, we wish to test that the distribution of marks in Statistics of B.Sc. students in college A and college B is same. If $F_1(x)$ and $F_2(x)$ are distribution functions of the marks in college A and college B respectively then our claim is $F_1(x) = F_2(x)$ and its complement is $F_1(x) \neq F_2(x)$. Since the claim contains the equality sign so we can take claim as the null hypothesis and complement as the alternative hypothesis. Thus,

$$H_0 : F_1(x) = F_2(x) \quad \text{for all values of x}$$

$$H_1 : F_1(x) \neq F_2(x) \quad \text{for at least one value of x}$$

So the test statistic is given by

$$D = \sup_x \left| S_1(x) - S_2(x) \right|$$

where, $S_1(x)$ and $S_2(x)$ are the empirical distribution functions of samples drawn from college A and college B respectively which are defined as:

$$S_1(x) = \frac{\text{number of students of college A whose marks} \le x}{n_1}$$

$$S_2(x) = \frac{\text{number of students of college B whose marks} \le x}{n_2}$$

Calculation for $\left| S_1(x) - S_2(x) \right|$:

Marks	Frequency (College A)	C.F. (College A)	$S_A(x)$	Frequency (College B)	C. F. (College B)	$S_B(x)$	$\left\|S_A(x) - S_B(x)\right\|$
0-10	2	2	0.0455	1	1	0.0278	0.0177
10-20	2	4	0.0909	1	2	0.0556	0.0354
20-30	4	8	0.1818	2	4	0.1111	0.0707
30-40	6	14	0.3182	5	9	0.2500	0.0682
40-50	3	17	0.3864	7	16	0.4444	0.0581
50-60	3	20	0.4545	3	19	0.5278	0.0732
60-70	4	24	0.5455	3	22	0.6111	0.0657
70-80	8	32	0.7273	2	24	0.6667	0.0606
80-90	7	39	0.8864	6	30	0.8333	0.0530
90-100	5	44	1	6	36	1	0
Total	$n_1 = 44$			$n_2 = 36$			

From the above calculation, we have

$$D = \sup_x \left| S_1(x) - S_2(x) \right|$$

$$= 0.0732$$

Since $n_1 = 44 > 16$ so we can calculate the critical value of test statistic for unequal sample sizes $n_1 = 44$ and $n_2 = 36$ at 1% level of significance by the formula

$$D_{n,\alpha} = 1.63\sqrt{\frac{n_1 + n_2}{n_1 n_2}} = 1.63\sqrt{\frac{44 + 36}{44 \times 36}} = 0.366$$

Since calculated value of test statistic D (= 0.0732) is less than critical value (= 0.366) so we do not reject the null hypothesis i.e. we support the claim at 1% level of significance.

Thus, we conclude that the samples fail to provide us sufficient evidence against the claim so we may assume that the distribution of marks in Statistics of B.Sc. students in college A and college B are same.

Now, you can try the following exercises.

E11) What is the main difference between Mann-Whitney U test and Kolmogorov-Smirnov two-sample tests?

E12) The following are the grades in Mathematics of High school students taken randomly from two School I and School II:

Grades	A	A^+	B	B^+	C	C^+	D	D^+
Number of Students in School- I	2	2	4	6	3	3	0	0
Number of Students in School- II	1	1	2	3	4	2	4	3

Apply Kolmogorov -Smirnov test to examine that the performance in Mathematics of High school students in School-I and School-II is same at 5% level of significance.

We now end this unit by giving a summary of what we have covered in it.

14.6 SUMMARY

In this unit, we have discussed following points:

1. Need of two-sample non-parametric tests.
2. Which non-parametric two-sample test is appropriate for a particular situation.
3. The test which is used in place of paired t-test when assumption(s) of the test is (are) not fulfilled, that is, paired sign test and Wilcoxon matched-pair signed-rank tests.
4. The Mann-Whitney U test which is used when the assumption(s) of two sample t-test is (are) not fulfilled.
5. The Kolmogorov-Smirnor two-sample test for testing the hypothesis that two samples come from the populations having the same distribution.

14.7 SOLUTIONS / ANSWERS

E1) The main difference between paired t-test and paired sign test is that the paired t-test based on normality assumption where as sign test is not based on such assumption.

E2) Here, the given dada are in the form of before and after study but assumption of normality is not given so we cannot use paired t-test. So we go for paired sign test instead of paired t-test.

Here, we wish to test that IQ is improved due to course designed that means average score before the course is less than average score after the course. If $\tilde{\mu}_1$ and $\tilde{\mu}_2$ denote the average (median) score before and after the course so our claim is $\tilde{\mu}_1 < \tilde{\mu}_2$ and its complement is $\tilde{\mu}_1 \geq \tilde{\mu}_2$. Since complement contains the equality sign so we can take the complement as the null hypothesis and the claim as the alternative hypothesis. Thus,

$$H_0 : \tilde{\mu}_1 \geq \tilde{\mu}_2$$

$$H_1 : \tilde{\mu}_1 < \tilde{\mu}_2 \qquad [\text{course improves the IQ}]$$

Since the alternative hypothesis is left-tailed so the test is left-tailed test and the test statistic (S) is the number of plus signs (S^+).

Calculation for S:

S. No	IQ Before (X)	IQ After (Y)	Sign of Difference (X – Y)
1	95	99	–
2	112	115	–
3	128	133	–
4	96	96	Tie
5	97	99	–
6	117	120	–
7	105	106	–
8	95	106	–
9	99	100	–
10	86	89	–
11	75	90	–
12	82	85	–

From the above calculation, we have

n = number of non-zero difference = 11

S^+ = number of plus signs = 0

S^- = number of minus signs = 11

Therefore, the value of the test statistic is

$$S = \text{number of plus signs } (S^+) = 0$$

To take the decision about the null hypothesis, we determine p-value with the help of **Table I** given in Appendix at the end of this block. Here, n = 11, p = 0.5 and r = 0. Thus, we have

$$\text{p-value} = P[S \le 0] = 0.0005$$

Since p-value = 0.0005< 0.05 (= α) so we reject the null hypothesis and support the alternative hypothesis i.e. we support the claim at 5% level of significance.

Thus, we conclude that the samples fail to provide us sufficient evidence against the claim so we may assume that IQ of students is improved due to course.

E3) Here, the ratings are given in the form of pairs so it can be considered as the case of dependent samples. But assumption of normality is not given also the given data are in the form of ranks so we cannot use the paired t-test in this case. So we go for paired sign test.

Here, we want to test that the product A is favoured over product B that means average (median) rating consumption of product A is greater than B. If $\tilde{\mu}_1$ and $\tilde{\mu}_2$ denote the average (median) rating consumption of product A and product B respectively then our claim is $\tilde{\mu}_1 > \tilde{\mu}_2$ and its complement is $\tilde{\mu}_1 \le \tilde{\mu}_2$. Since complement contains the equality sign so we can take the complement as the null hypothesis and the claim as the alternative hypothesis. Thus,

$$H_0 : \tilde{\mu}_1 \le \tilde{\mu}_2$$

$$H_1 : \tilde{\mu}_1 > \tilde{\mu}_2 \qquad \left[\text{the product A is favoured over product B }\right]$$

Since the alternative hypothesis is right-tailed so the test is right-tailed test so the test statistic (S) is the number of minus signs (S^-).

Calculation for S:

S. No	Favour of Product A (X)	Favour of Product B (Y)	Sign of Difference (X – Y)
1	5	2	+
2	2	3	–
3	4	1	+
4	4	2	+
5	5	3	+
6	4	4	Tie
7	5	4	+
8	5	1	+
9	2	4	–
10	3	2	+
11	3	2	+
12	4	3	+
13	3	3	Tie
14	3	4	–
15	2	3	–

From the above calculation, we have

n = number of non-zero differences = 13

S^+ = number of plus signs = 9

S^- = number of minus signs = 4

Therefore, value of test statistic is

S = number of minus signs (S^-) = 4

To take the decision about the null hypothesis, we determine p-value with the help of **Table I** given in Appendix at the end of this block. Here, n = 13, p = 0.5 and r = 4. Thus, we have

$$\text{p-value} = P\left[S \le 4\right] = 0.1334$$

Since p-value = 0.1334 > 0.01(= α) so we do not reject the null hypothesis and reject the alternative hypothesis i.e. we reject the claim at 1% level of significance.

Thus, we conclude that the samples provide us sufficient evidence against the claim so product A is not favoured over product B.

E4) Wilcoxon matched-pair singed-rank test is based on the assumption that the distribution of differences is symmetric whereas paired sign test does not require this assumption.

E5) It is case of two dependent populations and the assumption of normality for the differences is not given so we will not go for paired t-test. Also measurement of observations is available in interval scale so we will go for Wilcoxon matched pair signed-rank test.

Here, we wish to test that the students have benefited by coaching that means average score before the coaching is less than average score after the coaching. If $\tilde{\mu}_1$ and $\tilde{\mu}_2$ denote the average (median) score before and after the coaching then our claim is $\tilde{\mu}_1 < \tilde{\mu}_2$ and its complement is $\tilde{\mu}_1 \geq \tilde{\mu}_2$. Since complement contains the equality sign so we can take the complement as the null hypothesis and the claim as the alternative hypothesis. Thus,

$$H_0 : \tilde{\mu}_1 \geq \tilde{\mu}_2$$

$$H_1 : \tilde{\mu}_1 < \tilde{\mu}_2 \quad [\text{students have benefited by coaching}]$$

Since the alternative hypothesis is left-tailed so the test is left-tailed test.

Since in this exercise the increment in marks is given that means we are given that the difference of after on before, that is, **Y − X** instead of **X – Y** therefore, for testing the null hypothesis, the test statistic is the sum of negative ranks (T^-) instead of sum of positive ranks (T^+).

Calculation for T:

Roll No.	**Difference (d)**	**Absolute Difference $\lvert d \rvert$**	**Rank of $\lvert d \rvert$**	**Signed Rank**
1	6	6	6.5	6.5
2	−2	2	1.5	−1.5
3	8	8	8	8
4	−4	4	3.5	−3.5
5	10	10	9	9
6	2	2	1.5	1.5
7	5	5	5	5
8	−4	4	3.5	−3.5
9	6	6	6.5	6.5
10	Tie	---	---	---

From the above calculations, we have

$T^- = 8.5$

n = number of non-zero d_i's = 9

So the value of test statistic(T) is $T^- = 8.5$

The critical (tabulated) value of test statistic for one-tailed test corresponding to n = 9 at 5% level of significance is 9.

Since calculated value of test statistic T (= 8.5) is less than the critical value (= 9) so we reject the null hypothesis and support the alternative hypothesis i.e. we support the claim at 5% level of significance.

Thus, we conclude that the sample fails to provide us sufficient evidence against the claim so we may assume that the coaching is effective.

E6) Here, the data are given in the form if pairs so it can be considered as the case of dependent samples. But assumption of normality is not given also the given data are in the form of ranks so we cannot use the paired t-test in this case. So we go for Wilcoxon matched-pair signed-rank test.

Here, we wish to test that the product A and B are equally soft. If $\tilde{\mu}_1$ and $\tilde{\mu}_2$ denote the average rank of softness of products A and B respectively then our claim is $\tilde{\mu}_1 = \tilde{\mu}_2$ and it complement is $\tilde{\mu}_1 \neq \tilde{\mu}_2$. Since claim contains the equality sign so we can take the claim as the null hypothesis and the complement as the alternative hypothesis. Thus,

$$H_0 : \tilde{\mu}_1 = \tilde{\mu}_2 \text{ and } H_1 : \tilde{\mu}_1 \neq \tilde{\mu}_2$$

Since the alternative hypothesis is two-tailed so the test is two-tailed test so the test statistic is the minimum of the sum of positive ranks(T^+) and negative ranks (T^-), that is,

$$T = \min\left\{T^+, T^-\right\}$$

Calculation for T:

Judge	Product A	Product B	Difference (d)	Absolute Difference $\lvert d \rvert$	Rank of $\lvert d \rvert$	Signed Rank
1	4	5	–1	1	2	–2
2	6	4	2	2	5	5
3	8	5	3	3	7.5	7.5
4	9	8	1	1	2	2
5	4	1	3	3	7.5	7.5
6	7	9	–2	2	5	–5
7	6	2	4	4	9	9
8	5	3	2	2	5	5
9	8	2	6	6	10	10
10	6	7	–1	1	2	–2

From the above calculations, we have

$T^+ = 46$ and $T^- = 9$

n = number of non-zero d_i's = 10

Putting the values in test statistic, we have

$$T = \min\left\{T^+, T^-\right\} = \min\left\{46, 9\right\} = 9$$

The critical value of test statistic for two-tailed test corresponding n = 10 at 1% level of significance is 4.

Since calculated value of test statistic T (= 9) is greater than the critical value (= 4) so we do not reject the null hypothesis i.e. we support the claim at 1% level of significance.

Thus, we conclude that the sample fail to provide us sufficient evidence against the claim so we may assume that the products A and B are equally soft.

E7) Two-sample t-test is based on the assumptions that populations under study are normally distributed and the observations are measured in at least of an interval scale whereas Mann-Whitney U test is non required such assumptions.

E8) Wilcoxon matched-pair singed-rank test is used when observations are paired whereas the Mann-Whitney U test is used when two independent samples are randomly drawn two populations.

E9) It is case of two independent populations and the assumption of normality of both the populations is not given so we cannot use t-test in this case. So we go for Mann-Whitney U test.

Here, we want to test that village boys are more fit than town boys that means average (median) score of village boys is greater than average (median) score of town boys. If $\tilde{\mu}_1$ and $\tilde{\mu}_2$ denote the average (median) score of the village boys and town boys respectively then our claim is $\tilde{\mu}_1 > \tilde{\mu}_2$ and its complement is $\tilde{\mu}_1 \le \tilde{\mu}_2$. Since complement contains the equality sign so we can take the complement as the null hypothesis and the claim as the alternative hypothesis. Thus,

$$H_0 : \tilde{\mu}_1 \le \tilde{\mu}_2$$

$$H_1 : \tilde{\mu}_1 > \tilde{\mu}_2 \quad [\text{village boys are more fit than town boys}]$$

To perform the Mann-Whitney U test, we combine the observations of both samples and for convenient, we arrange the observations in ascending order. Then rank all these combined observations from smallest to largest as:

Village Boys(V)	Ranks	Town Boys (T)	Ranks
4.5	5	2.7	1
5.6	6.5	2.8	2
6.5	9.5	3.2	3
7.2	12	3.6	4
8.2	15	5.6	6.5
9.0	16	6.1	8
10.6	17	6.5	9.5
11.4	18	6.7	11
12.4	20	7.9	13.5
12.9	23	7.9	13.5
15.7	23	11.8	19
16.2	25	12.6	21
		12.7	22
Total	190		

Here, n_1 =12 and n_2 = 13. Since n_1 is less than n_2 therefore we take sum of ranks corresponding to village boys.

Since we consider the first sample so test statistic U is given by

$$U = S - \frac{n_1(n_1+1)}{2} = 190 - \frac{12 \times (12+1)}{2} = 112$$

The upper critical value of test statistic for right-tailed corresponding to n_1 = 12 and n_2 = 13 at 5% level of significance is $U_{U,\alpha} = U_{U,\,0.01} = 198.$

Since calculated value of test statistic (=112) is less than upper critical value (= 198) so we do not reject the null hypothesis and reject the alternative hypothesis i.e. we support the claim at 5% level of significance.

Thus, we conclude that the samples provide us sufficient evidence against the claim so village boys are not more fit than town boys.

E10) It is case of two independent populations and the assumption of normality of both the populations is not given so we cannot use two sample t-test in this case. So we go for Mann-Whitney U test.

Here, we want to test that average life of two brands of batteries is same. If $\tilde{\mu}_1$ and $\tilde{\mu}_2$ denote the average (median) life of batteries of brands A and B respectively then our claim is $\tilde{\mu}_1 = \tilde{\mu}_2$ and it complement is $\tilde{\mu}_1 \neq \tilde{\mu}_2$. Since claim contains the equality sign so we can take the claim as the null hypothesis and the complement as the alternative hypothesis. Thus

$$H_0 : \tilde{\mu}_1 = \tilde{\mu}_2 \; [\text{average life of two brands of batteries is same}]$$

$$H_1 : \tilde{\mu}_1 \neq \tilde{\mu}_2 \; [\text{average life of two brands of batteries is not same}]$$

To perform the Mann-Whitney U test, we combine the observations of both samples and for convenient, we arrange the observations in ascending order. Then rank all these combined observations from smallest to largest as:

Brand A	**Ranks**	**Brand B**	**Ranks**
30	1.5	35	4
30	1.5	40	8.5
35	4	45	12.5
35	4	50	16
40	8.5	50	16
40	8.5	50	16
40	8.5	50	16
40	8.5	55	20
40	8.5	55	20
45	12.5	60	23
50	16	60	23
55	20	60	23
Total	102		

Here, $n_1 = n_2$ so we can take any the sum of ranks(S) of first sample or second sample. We consider first sample. Therefore, the test statistic U is given by

$$U = S - \frac{n_1(n_1+1)}{2} = 102 - \frac{12 \times (12+1)}{2} = 24$$

The critical values of test statistic for two-tailed test corresponding $n_1 = n_2 = 12$ at 1% level of significance are $U_{L,\alpha/2} = U_{L,\,0.005} = 106$ and $U_{U,\alpha/2} = U_{U,\,0.005} = 194$.

Since calculated value of test statistic U (= 24) is less than both critical values (= 106 and 194) so we reject the null hypothesis i.e. we reject the claim at 1% level of significance.

Thus, we conclude that the samples provide us sufficient evidence against the claim so average life of two brands of batteries is not same.

E11) Mann-Whitney U test is used to test the hypothesis that two independent samples come from the populations follow distributions that are identical with respect to location whereas Kolmogorov-Smirnov two-sample test is sensitive to differences of all types that may exist between two distributions that is, location, dispersion, skewness, etc.

E12) Here, we want to test that performance of students in School-I and School-II are same. If $F_1(x)$ and $F_2(x)$ are distribution functions of grades in Mathematics of High school students in School-I and School-II then our claim is $F_1(x) = F_2(x)$ and its complement is $F_1(x) \neq F_2(x)$. Since the claim contains the equality sign so we can take claim as the null hypothesis and complement as the alternative hypothesis. Thus,

$$H_0 : F_1(x) = F_2(x) \qquad \text{for all values of x}$$

$$H_1 : F_1(x) \neq F_2(x) \qquad \text{for at least one value of x}$$

For testing the null hypothesis, the test statistic is given by

$$D = \sup_x \left| S_1(x) - S_2(x) \right|$$

where, $S_1(x)$ and $S_2(x)$ are the empirical cumulative distribution functions of samples drawn from School-I and School-II respectively which are defined as:

$$S_1(x) = \frac{\text{number of students of School-I whose grade} \leq x}{n_1}$$

and

$$S_2(x) = \frac{\text{number of students of School-II whose grade} \leq x}{n_2}$$

Calculation for $\left|S_1(x) - S_2(x)\right|$:

Grade	**Frequency (School-I)**	**C.F. (School-I)**	**$S_A(x)$**	**Frequency (School-II)**	**C. F. (School-II)**	**$S_B(x)$**	**$\left|S_A(x) - S_B(x)\right|$**
A	2	2	0.10	1	1	0.05	0.05
A^+	2	4	0.20	1	2	0.10	0.10
B	4	8	0.40	2	4	0.2	0.20
B^+	6	14	0.70	3	7	0.35	0.35
C	3	17	0.85	4	11	0.55	0.30
C^+	3	20	1.00	2	13	0.65	0.35
D	0	20	1.00	4	17	0.85	0.15
D^+	0	20	1.00	3	20	1	0.00
Total	$n_1 = 20$			$n_2 = 20$			

From the above calculation, we have

$$D = \sup_x \left| S_1(x) - S_2(x) \right| = 0.35$$

The critical value of test statistic for equal sample sizes $n_1 = n_2 = 20$ at 5% level of significance is 8/20 = 0.4.

Since calculated value of test statistic (= 0.35) is less than critical value (= 0.4) so we do not reject the null hypothesis i.e. we support the claim at 5% level of significance.

Thus, we conclude that the samples fail to provide us sufficient evidence against the claim so we may assume that the performance of students in School-I and School-II is same.