11.1 Sign Test – Paired Samples.

To test the value of one variable against that of another or paired dependent samples, the Sign test can be used[61]. The null hypothesis will be that there is no difference between the variables. That signifies that the variables come from the same population. Only the sign of the difference is used. The smaller number of plus or minus signs becomes the test value. If the sums of the plus and minus signs were +6 and -5, then the test value would be 5 (the smaller absolute value is used). Comparing the test value to the critical value, *reject or fail to accept the null hypothesis if the test value is equal to or less than the critical value*. For small sample sizes, use **Appendix C**, **Table C-8,** for the Sign Test; otherwise, the z-distribution table is used. The formula to determine the critical value for larger samples (over 25) is $z = \frac{(X + 0.5) - (n/2)}{\sqrt{n}/2}$. The X value is the smaller number of plus or minus signs and n is the sample size. If there is the same value, then that counts as a zero (0) and no plus or minus value is given. Also, the Sign test can be used for a single sample against a theorized median; however, the paired comparison test is being shown here in **Example 11.1**.

Example 11.1 Sign Test – Paired Samples

A golf instructor wanted to determine if his golf lessons had a beneficial impact on the students to whom he gave lessons. He sampled 10 players; their known golf averages were compared to after-lesson round-averages over a two-month period. **From the data below, can the instructor conclude at α = .05 that his lessons improved the golf averages?**

Golfer	Before Lessons	After Lessons	Golfer	Before Lessons	After Lessons
1	85	84	**6**	90	91
2	95	92	**7**	112	110
3	110	100	**8**	89	90
4	84	84	**9**	84	84
5	115	112	**10**	120	112

Solution:

Step 1. **State the hypotheses**.
H_0: The golf lessons will not improve the average golf scores.
H_A: The golf lessons will improve the average golf scores.

Step 2. **Determine the critical value**. Use the Sign Test Table (**Appendix C, Table C-8, one-tail**) and the lowest positive or negative sign number at α = .05 (do not count zeros) for the test value.

[61] The t-test is used for normally distributed samples. The Sign test is used for non-normal distributions.

Example 11.1 Sign Test – Paired Samples

Golfer	Before Lessons	After Lessons	Sign of Diff.	Golfer	Before Lessons	After Lessons	Sign of Diff.
1	85	84	+	**6**	90	91	+
2	95	92	+	**7**	112	110	+
3	110	100	+	**8**	89	90	**-**
4	84	84	**0**	**9**	84	84	**0**
5	115	112	+	**10**	120	112	+

Step 3. **Determine the test value** and **compare it to the critical value**. Lowest sign is one negative against critical value of one: 7 (+) SIGNS AND 1 (-) NEGATIVE SIGN. n = 8 @ α = .05 (one-tail) gives the critical value = 1. If the smaller test value equals or is less than 1, reject the null hypothesis.

Step 4. **Make the decision**. **Since the lowest sign value equals the critical value, reject the null hypothesis that the golf lessons will not improve the average golf scores.** Suggest more lessons. Comments can be made that one golfer did not equal or lower the golf average and two golfers had no change to their averages. However, six golfers did make an improvement.

11.2 Wilcoxon Tests.

The Sign test does not consider the magnitude of the data, only if a datum is over or under a median or paired datum. By using ranks, the Wilcoxon Rank Sum and the Wilcoxon Signed-Rank tests can be used. The former is for independent samples and the latter for dependent samples. Again, these are for non-parametric distributions while the *t*- and *z*-test are for parametric distributions. For a null hypothesis, there should be no difference in the distributions (two-tailed). A large difference means the alternate hypothesis of a difference is accepted or a directional finding is found. For example, if one sample is measured against the median and ordinal data are available, this test is useful. Measuring two samples against one another is a paired comparison and will be shown below. The standard alpha values are generally used (.01 and .05).

11.2.1 Wilcoxon Rank Sum Test.

For independent samples larger than or equal to 10, the Wilcoxon Rank Sum[62] test can be used. The formula for this test is $z = \frac{R - \mu_R}{\sigma_R}$ where $\mu_R = \frac{n_1(n_1 + n_2 + 1)}{2}$ and

$\sigma_R = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$. Also, R is the sum of ranks for the smaller sample size (n_1). The larger sample size is n_2. Both samples must be equal to or larger than 10. If the same

[62] Also called the Mann-Whitney Rank Sum Test.

number, it doesn't matter which is declared n_1. **Example 11.2** shows the *independent samples* test.

Example 11.2 Wilcoxon Rank Sum Test

Two independent samples of 10 male and 10 female students were asked to rate their value (on a 100 point scale) of having to take two years of general academic courses versus focusing more on their specialty field courses and programs. The higher the rating, the greater it was to value taking more specialty courses and not take so many other academic courses. It was theorized that there would be a difference in the opinions based on gender. **At α = .05, is there a difference based upon gender?**

Males	90 92 88 78 82 84 90 86 78 86	Mean = 85.4
Females	88 94 96 90 86 87 94 96 89 95	Mean = 91.5

Solution:

Step 1. **State the hypotheses**.

H_0: There is no difference in gender preferences of having to take general academic courses in a specialty degree program.

H_A: There is a difference in gender preferences of having to take general academic courses in a specialty degree program.

Step 2. **Determine the critical *z*-value** from **Appendix C, Table C-2**[63]**,** at $\alpha = .05$ for a two-tailed test. Since this is at $\alpha = .05$ for two tails, each side of the distribution has .025 or a *z*-value of ± 1.96.

Step 3. **Compute the test value**.

a. **Tabulate the data in order, lowest to highest, and rank each value**.

Value	78	78	82	84	86	86	86	87	88	88
	89	90	90	90	92	94	94	95	96	96
Gender	M	M	M	M	M	M	F	F	M	F
	F	M	M	F	M	F	F	F	F	F
Rank	1.5	1.5	3	4	6	6	6	9	10.5	10.5
	12	14	14	14	16	17.5	17.5	19	19.5	19.5

b. **Sum the ranks of the group with the smaller sample size**. Both groups are equal in size. Males are chosen as n_1.

R = 1.5 + 1.5 + 3 + 4 + 6 + 6 + 10.5 + 14 + 14 + 16 = 81

c. **Use the formulas to find the test value**.

$$\mu_R = \frac{n_1(n_1+n_2+1)}{2} = \frac{10(10+10+1)}{2} = \frac{210}{2} = 105$$

$$\sigma_R = \sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}} = \sqrt{\frac{(10)(10)(10+10+1)}{12}} = \sqrt{175} = 13.23$$

$$z = \frac{R-\mu_R}{\sigma_R} = \frac{81-105}{13.23} = \mathbf{-1.814}$$

Step 4. **Decision: -1.96 < -1.814 < +1.96. Accept the null hypothesis. There is no significant difference in gender preferences** of having to take general academic courses in a specialty degree program.

[63] Use the *z*-table of critical values for the Wilcoxon Rank Sum Test.

11.2.2 Wilcoxon Signed-Rank Test.

The only assumptions for the Wilcoxon Signed-Rank test are that (a) the populations have a symmetrical distribution, (b) the samples are dependent, and (c) it does not require normality. If the sample size is ≥ 30, the following formula is used to compute the observed *z*-value: $z = \dfrac{w_s - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$. **Example 11.3** shows the *dependent samples* test. As the sample size is < 30, the solution strategy is simplified: (a) Determine the Wilcoxon Signed-Rank critical value from **Appendix C, Table C-9**; (b) compute the positive and negative rank sums as shown; (c) select the smaller absolute value; (d) set that value equal to (w_s); and then (e) compare the observed w_s to the critical value.

Example 11.3 Wilcoxon Signed-Rank Test

In a statistics class, ten students were sampled based upon their different course requirements. For the first part of the course, each student studied individually while team study requirements were established after the midterm exam. The instructor wanted to determine if a difference in their course study requirement was noted. **At alpha .05, can it be concluded that there was a difference in their exams marks based upon the course requirement?** (dependent samples)

Student	1	2	3	4	5	6	7	8	9	10
Midterm	90	92	84	78	90	88	94	82	77	89
Final	88	90	87	90	95	85	92	92	86	88

Solution:
Step 1. **State the hypotheses**.
H_0: There is no difference in exam grades based upon the course study requirements.
H_A: There is a difference in exam grades based upon the course study requirements.

Step 2. **Determine the critical value** from the Wilcoxon Signed-Rank table (**Appendix C, Table C-9**). Since n = 10 and α = .05 for this two-tailed test, the **critical value is 8**.

Step 3. **The test value is found by developing a table**.

Student	Midterm	Final	Difference	Absolute Value	Rank	Signed Rank	Positive rank sum	Negative rank sum
1	90	88	-2	2	2.5	-2.5		-2.5
2	92	90	-2	2	2.5	-2.5		-2.5
3	84	87	+3	3	4.5	+4.5	+4.5	
4	78	90	+12	12	10	+10	+10	
5	90	95	+5	5	6	+6	+6	
6	88	85	-3	3	4.5	-4.5		-4.5
7	88	94	+6	6	7	+7	+7	
8	82	92	+10	10	9	+9	+9	
9	77	86	+9	9	8	+8	+8	
10	89	88	-1	1	1	-1		-1
						Σ =	**+44.5**	**-10.5**

Example 11.3 Wilcoxon Signed-Rank Test	
	Select the smaller of the absolute values of the sums: $|-10.5| = 10.5$. Use this for the test value (w_s) as sample size < 30. w_s = **10.5.**
Step 4.	**Decide whether the test value is less than or equal to the critical value of 8.** **Here 10.5 > 8 so fail to accept the null.** **Accept the alternate: The team study effort made a positive difference in the grading.**

11.3 Kruskal-Wallis Test.

When using known population parameters, the ***F*-test** for three or more means is preferred. When the population parameters are unknown, the Kruskal-Wallis test for three or more means is used. This is also called the ***H* test**. Each sample size must be five or more and ranking is used. The test is always right-tailed with degrees of freedom (df) equal to (k – 1; k = number of groups). To obtain the critical value, the Chi Square Table is used (**Appendix C, Table C-4**).

The formula for this test is $H = \frac{12}{N(N+1)}\left(\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \cdots + \frac{R_k^2}{n_k}\right) - 3(N+1)$.

N equals the sum of all the sample sizes in k groups ($n_1 + n_2 + \ldots n_k$). R_1 equals the sum of ranks in sample one. R_2 equals the sum of ranks in sample 2 and so forth for each sample. The $n_1 + n_2 + \ldots n_k$ values are the number of samples in the 1st, 2nd, …kth groups.

Example 11.4 Kruskal-Wallis Test

A specific medical treatment program was developed by the U.S. Medical Services (Air Force, Army, & Navy) for treating mental conditions of personnel who served in hostile environments. Ten randomly selected personnel from each Service were evaluated using the same testing procedures and each person was assigned a score based upon a 50-point scale. **At α = .05, is there a sufficient evidence to declare the treatment results are different among the Services?**

Air Force	42, 46, 45, 42, 40, 48, 46, 49, 41, 44
Army	38, 40, 46, 44, 39, 47, 42, 44, 45, 41
Navy	48, 44, 46, 40, 39, 42, 46, 47, 48, 45

Step 1. **State the hypotheses.**
H_0: There is no difference in the results of the medical treatment program among the three Service groups.
H_A: There is a difference in the results of the medical treatment program among the three Service groups.

Step 2. **Determine the critical value using the Chi Square distribution table**.
df = k-1 or 3 -1 = 2 α = .05 **Critical value = 5.991**

Step 3. **Compute the test value**. Arrange the data from lowest to highest and rank each value. Identify the Groups and sum each Group ranking.

Example 11.4 Kruskal-Wallis Test

Score	Group	Rank	Score	Group	Rank
38	A (Army)	1	44	N	15
39	N (Navy)	2.5	45	A	18
39	A	2.5	45	N	18
40	AF (Air Force)	5	45	AF	18
40	A	5	46	AF	22
40	N	5	46	AF	22
41	AF	7.5	46	A	22
41	A	7.5	46	N	22
42	AF	10.5	46	N	22
42	AF	10.5	47	A	25.5
42	A	10.5	47	N	25.5
42	N	10.5	48	AF	28
43	AF	13	48	N	28
44	A	15	48	N	28
44	A	15	49	AF	30

Sum each Group ranking.

Air Force	5	7.5	10.5	10.5	13	18	22	22	28	30	= 166.5
Army	1	2.5	5	7.5	10.5	15	15	18	22	25.5	= 122
Navy	2.5	5	10.5	15	18	22	22	25.5	28	28	= 176.5

Apply the formula: $H = \frac{12}{N(N+1)}\left(\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \cdots + \frac{R_k^2}{n_k}\right) - 3(N+1)$

$$H = \frac{12}{30(30+1)}\left(\frac{166.5^2}{10} + \frac{122^2}{10} + \frac{176.5^2}{10}\right) - 3(30+1)$$

$$H = \frac{12}{930}\left(\frac{27{,}722.25}{10} + \frac{14{,}884}{10} + \frac{31{,}152.25}{10}\right) - 93$$

H = 0.0129 (7375.8) – 93 = 95.148 – 93 = **2.148**

Step 4. **Make the Decision and summarize the results**. **2.148 < 5.991**
Accept the null hypothesis that there is no difference in the results of the medical treatment program among the three Service groups at alpha .05.
A review of the rankings can indicate that the Navy had the highest rankings and the Army had the lowest. Further sampling could be conducted.

11.4 Mann-Whitney U Test.

When there are two independent groups and the dependent variable is ordinal or continuous, and the distribution is not normal, the Mann-Whitney U test can be applied. An example would be gender difference attitudes over working conditions. The independent variable would be gender (male and female) and the dependent variable would be attitudes toward working conditions. When the distribution would be normal, the independent *t*-test would be applicable.

As in each test, various assumptions are necessary. The first assumption is that the dependent variable is measurable at the ordinal or continuous level. A Likert scale that has various marking levels; for example, Strongly Agree, Agree, Strongly Disagree, and so forth, can be used to rank the categories. Continuous markings are used for time, weight, and other categories.

The second assumption relates to having two independent groups. The male and female groups meet this category. Independent observation indicating no relationships is another assumption. For example, in each group above (male and female), different participants respond. Lastly, the non-normal distribution is assumed. Any samples under consideration do not have to be numerically the same.

The data obtained is combined and ranking is done. The Mann-Whitney U formula is $U_1 = n_1 \bullet n_2 + [n_1(n_1 + 1) / 2] - R_1$. The R_1 indicates the sum of the ranks for the first sample. The n_1 and n_2 are the first and second sample sizes. The formula for the second sample is $U_2 = n_1 \bullet n_2 + [n_2(n_2 + 1) / 2] - R_2$. The Mann-Whitney Table of critical values (**Appendix C, Table C-11**) is used to compare the lowest U value determined. The null hypothesis is that the populations that the two samples came from have identical median values. If the U value at $\alpha = .05$ or .01 is less than or equal to the critical table value, the groups are not equal. The test can also be one-tailed. **Example 11.5** shows this application. Other formulas for the Mann-Whitney U test are also available, but the results should be similar.

Example 11.5 Mann-Whitney U Test

At Cambridge Commodities, the Managing Director wanted to determine if the attitudes of the males and females was identical based upon the provision of new showers, kitchens, and gym facilities he installed in the new company building. Two independent samples were taken based upon a scoring system of 50 points; **can it be concluded at alpha .05 that the attitudes of the two groups were identical median values***?* Ranking has been completed.

Males				Females			
Score	**Rank**	**Score**	**Rank**	**Score**	**Rank**	**Score**	**Rank**
48	17.5	43	9	42	7.5	46	12.5
46	12.5	41	5.5	41	5.5	42	7.5
46	12.5	47	15.5	49	18.5	40	3
40	3	45	10	48	17.5	49	19.5
40	3			47	15.5	46	12.5
				39	1		

Example 11.5 Mann-Whitney U Test	
Step 1.	**State the hypotheses**. H_0: The medians of the males and females have identical values in regards to attitudes of the new equipment in the company. H_A: The medians of the males and females do *not* have identical values in regards to attitudes of the new equipment in the company.
Step 2.	**Sum the group ranks**. Males = 17.5 + 12.5 + 12.5 + 3 + 3 + 9 + 5.5 + 15.5 + 10 = 88.5 = R_1 Females = 7.5 + 5.5 + 18.5 + 17.5 + 15.5 + 1 + 12.5 + 7.5 + 3 + 19.5 + 12.5 = 120.5 = R_2
Step 3.	**Substitute into the formulas**. $U_1 = n_1 \bullet n_2 + [n_1(n_1 + 1) / 2] - R_1$; $U_2 = n_1 \bullet n_2 + [n_2(n_2 + 1) / 2] - R_2$. $U_1 = 9\ (11) + [\ 9(9 + 1) / 2] - 88.5 = 99 + 45 - 88.5 =$ **55.5 =** $\mathbf{U_1}$ $U_2 = 9\ (11) + [11(11 + 1) / 2] - 120.5 = 99 + 66 - 120.5 =$ **44.5 =** $\mathbf{U_2}$
Step 4.	**Compare the *lower* computed value (44.5) to the critical** value using Mann-Whitney U table (A**ppendix C, Table C-11**) at $\alpha = .05$. **Critical value = 23 at .05, two-tailed; therefore 44.5 > 23**. To reject the null, the determined value must be ≤ 23; therefore, **accept the null that the medians of the two groups are the same**.

Summary

Various non-parametric tests (Sign test – Paired Samples, Wilcoxon Rank Sum Test, Wilcoxon Signed-Rank Test, Kruskal-Wallis, and Mann-Whitney U test) are presented. Ranking is used in many of these tests. The advantages of using a non-parametric test include (a) using nominal or ordinal data, (b) not needing to have the population boundaries, (c) ease of use and understanding, and (d) not needing to have a normal distribution. To test the value of one variable to that of another or paired dependent samples, the Sign test can be used. The Sign test does not consider the magnitude of the data, only if a datum is over or under a median or paired datum. By using ranks, the Wilcoxon rank sum and the Wilcoxon signed-rank tests can be used. The former is for independent samples and the latter for dependent samples. When the population parameters are unknown, the Kruskal-Wallis test for three or more means is used. This is also called the *H* test. The U test is the Mann-Whitney U and is used to compare two independent samples using ordinal or continuous data. See **Figures 11.1** and **11.2** for an overview of this chapter.