4

COMPARING TWO UNRELATED SAMPLES: THE MANN–WHITNEY U-TEST

4.1 OBJECTIVES

In this chapter, you will learn the following items.

- How to compute the Mann–Whitney U-test.
- How to perform the Mann–Whitney U-test using SPSS.
- How to construct a median confidence interval based on the difference between two independent samples.

4.2 INTRODUCTION

Suppose a teacher wants to know if his first-period's early class time has been reducing student performance. To test his idea, he compares the final exam scores of students in his first-period class with those in his fourth-period class. In this example, each score from one class period is independent of, or unrelated to, the other class period.

The Mann–Whitney U-test is a nonparametric statistical procedure for comparing two samples that are independent, or not related. The Wilcoxon rank sum test is a similar nonparametric test to the Mann–Whitney U-test. The parametric equivalent to these tests is the t-test for independent samples.

In this chapter, we will describe how to perform and interpret a Mann–Whitney U-test for both small samples and large samples. We will also explain how to perform

Nonparametric Statistics for Non-Statisticians, Gregory W. Corder and Dale I. Foreman

the procedure using SPSS. Finally, we offer varied examples of these nonparametric statistics from the literature.

4.3 COMPUTING THE MANN–WHITNEY *U*-TEST STATISTIC

The Mann–Whitney *U*-test is used to compare two unrelated, or independent, samples. The two samples are combined and rank ordered together. The strategy is to determine if the values from the two samples are randomly mixed in the rank ordering or if they are clustered at opposite ends when combined. A random rank order would mean that the two samples are not different, while a cluster of one sample's values would indicate a difference between them. In Figure 4.1, two sample comparisons illustrate this concept.

The scores in Comparison 1 are rank-ordered in clusters at opposite ends. This suggests that treatment X might be higher than treatment O.	Comparison 1 X X X O X X X X O O O O 1 2 3 4 5 6 7 8 9 10 11 12
The scores in Comparison 2 are spread along the entire distribution. This suggests that there is no clear difference between treatments.	Comparison 2 X O O X X O X O X O X X 1 2 3 4 5 6 7 8 9 10 11 12

FIGURE 4.1

Use Formula 4.1 to determine a Mann–Whitney *U*-test statistic for each of the two samples. The smaller of the two *U* statistics is the obtained value.

$$U_i = n_1 n_2 + \frac{n_i(n_i+1)}{2} - \sum R_i \tag{4.1}$$

where U_i is the test statistic for the sample of interest, n_i is the number of values from the sample of interest, n_1 is the number of values from the first sample, n_2 is the number of values from the second sample, and $\sum R_i$ is the sum of the ranks from the sample of interest.

After the *U* statistic is computed, it must be examined for significance. We may use a table of critical values (see Table B.4). However, if the number of values in each sample, n_i, exceeds those available from the table, then a large sample approximation may be performed. For large samples, compute a z-score and use a table with the normal distribution (see Table B.1) to obtain a critical region of z-scores. Formulas 4.2–4.4 are used to find the z-score of a Mann–Whitney U-test for large samples.

$$\bar{x}_U = \frac{n_1 n_2}{2} \tag{4.2}$$

where $\bar{x}_U$ is the mean, n_1 is the number of values from the first sample, and n_2 is the number of values from the second sample.

$$s_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} \tag{4.3}$$

where s_U is the standard deviation.

$$z^* = \frac{U_i - \bar{x}_U}{s_U} \tag{4.4}$$

where z^* is the z-score for a normal approximation of the data and U_i is the U statistic from the sample of interest.

At this point, the analysis is limited to identifying the presence or absence of a significant difference between the groups and does not describe the strength of the treatment. We can consider the effect size (ES) to determine the degree of association between the groups. We use Formula 4.5 to calculate the effect size.

$$\text{ES} = \frac{|z|}{\sqrt{n}} \tag{4.5}$$

where $|z|$ is the absolute value of the z-score and n is the total number of observations.

The effect size ranges from 0 to 1. Cohen (1988) defined the conventions for effect size as small = 0.10, medium = 0.30, and large = 0.50. (Correlation coefficient and effect size are both measures of association. See Chapter 7 concerning correlation for more information on Cohen's assignment of effect size's relative strength.)

4.3.1 Sample Mann–Whitney U-Test (Small Data Samples)

The following data were collected from a study comparing two methods being used to teach reading recovery in the fourth grade. Method 1 was a pullout program in which the children were taken out of the classroom for 30 min. a day, 4 days a week. Method 2 was a small group program in which children were taught in groups of four or five for 45 min. a day in the classroom, 4 days a week. The students were tested using a reading comprehension test after 4 weeks of the program. The test results are shown in Table 4.1.

1. *State the null and research hypotheses.*

   The null hypothesis, shown below, states that there is no tendency of the ranks of one method to be systematically higher or lower than those of the other.

TABLE 4.1

Method 1	Method 2
48	14
40	18
39	20
50	10
41	12
38	102
53	17

The hypothesis is stated in terms of comparison of distributions, not means. The research hypothesis, shown below, states that the ranks of one method are systematically higher or lower than those of the other. Our research hypothesis is a two-tailed, nondirectional hypothesis because it indicates a difference, but in no particular direction.

The null hypothesis is

H_0: There is no tendency for ranks of one method to be significantly higher (or lower) than those of the other.

The research hypothesis is

H_A: The ranks of one method are systematically higher (or lower) than those of the other.

2. *Set the level of risk (or the level of significance) associated with the null hypothesis.*

   The level of risk, also called an alpha (α), is frequently set at 0.05. We will use an alpha of 0.05 in our example. In other words, there is a 95% chance that any observed statistical difference will be real and not due to chance.

3. *Choose the appropriate test statistic.*

   The data are obtained from two independent, or unrelated, samples of fourth-grade children being taught reading. Both the small sample sizes and an existing outlier in the second sample violate our assumptions of normality. Since we are comparing two unrelated, or independent, samples, we will use the Mann–Whitney U-test.

4. *Compute the test statistic.*

   First, combine and rank both data samples together (see Table 4.2). Next, compute the sum of ranks for each method. Method 1 is $\sum R_1$ and Method 2 is $\sum R_2$. Using Table 4.2,

$$\begin{aligned}\sum R_1 &= 7+8+9+10+11+12+13\\ &= 70\end{aligned}$$

TABLE 4.2

Ordered Scores		
Rank	Score	Sample
1	10	Method 2
2	12	Method 2
3	14	Method 2
4	17	Method 2
5	18	Method 2
6	20	Method 2
7	38	Method 1
8	39	Method 1
9	40	Method 1
10	41	Method 1
11	48	Method 1
12	50	Method 1
13	53	Method 1
14	102	Method 2

and

$$\begin{aligned}\sum R_2 &= 1+2+3+4+5+6+14 \\ &= 35\end{aligned}$$

Now, compute the U value for each sample. For sample 1,

$$\begin{aligned}U_1 &= n_1 n_2 + \frac{n_1(n_1+1)}{2} - \sum R_1 \\ &= 7(7) + \frac{7(7+1)}{2} - 70 = 49 + 28 - 70 \\ &= 7\end{aligned}$$

and for sample 2,

$$\begin{aligned}U_2 &= n_1 n_2 + \frac{n_2(n_2+1)}{2} - \sum R_2 \\ &= 7(7) + \frac{7(7+1)}{2} - 35 = 49 + 28 - 35 \\ &= 42\end{aligned}$$

The Mann–Whitney U-test statistic is the smaller of U_1 and U_2. Therefore, $U = 7$.

5. *Determine the value needed for rejection of the null hypothesis using the appropriate table of critical values for the particular statistic.*

Since the sample sizes are small ($n < 20$), we use Table B.4, which lists the critical values for the Mann–Whitney U. The critical values are found in the table at the point for $n_1 = 7$ and $n_2 = 7$. We set $\alpha = 0.05$. The critical value for the Mann–Whitney U is 8. A calculated value that is less than or equal to 8 will lead us to reject our null hypothesis.

6. *Compare the obtained value to the critical value.*

   The critical value for rejecting the null hypothesis is 8 and the obtained value is $U = 7$. If the critical value equals or exceeds the obtained value, we must reject the null hypothesis. If instead, the critical value is less than the obtained value, we must not reject the null hypothesis. Since the critical value exceeds the obtained value, we must reject the null hypothesis.

7. *Interpret the results.*

   We rejected the null hypothesis, suggesting that a real difference exists between the two methods. In addition, since the sum of the ranks for Method 1 (ΣR_1) was larger than that for Method 2 (ΣR_2), we see that Method 1 had significantly higher scores.

8. *Reporting the results.*

   The reporting of results for the Mann–Whitney U-test should include such information as the sample sizes for each group, the U statistic, the p-value's relation to α, and the sums of ranks for each group.

   For this example, two methods were used to provide students with reading instruction. Method 1 involved a pullout program and Method 2 involved a small group program. Using the ranked reading comprehension test scores, the results indicated a significant difference between the two methods ($U = 7$, $n_1 = 7$, $n_2 = 7$, $p < 0.05$). The sum of ranks for Method 1 ($\Sigma R_1 = 70$) was larger than the sum of ranks for Method 2 ($\Sigma R_2 = 35$). Therefore, we can state that the data support the pullout program as a more effective reading program for teaching comprehension to fourth-grade children at this school.

4.3.2 Performing the Mann–Whitney *U*-Test Using SPSS

We will analyze the data from the above example using SPSS.

1. *Define your variables.*

   First, click the "Variable View" tab at the bottom of your screen. Then, type the names of your variables in the "Name" column. Unlike the Wilcoxon signed ranks test described in Chapter 2, you cannot simply enter each sample into a separate column to execute the Mann–Whitney U-test. You must use a grouping variable. As shown in Figure 4.2, the first variable is the grouping variable that we called "Method". The second variable that we called "Score" will have our actual values.

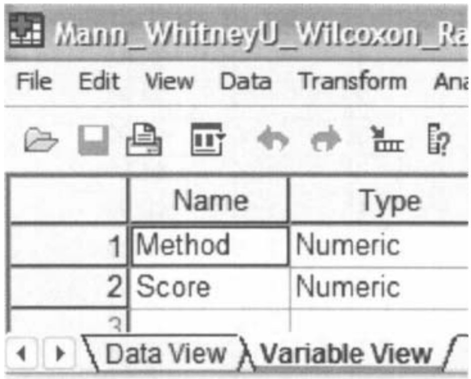


FIGURE 4.2

When establishing a grouping variable, it is often easiest to assign each group a whole number value. In our example, our groups are "Method 1" and "Method 2". Therefore, we must set our grouping variables for the variable "Method". First, we selected the "Values" column and clicked the gray square, as shown in Figure 4.3. Then, we set a value of 1 to equal "Method 1". Now, as soon as we click the "Add" button, we will have set "Method 2" equal to 2 based on the values we inserted above.

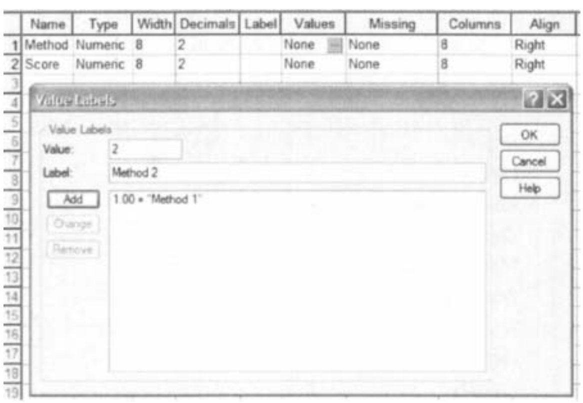
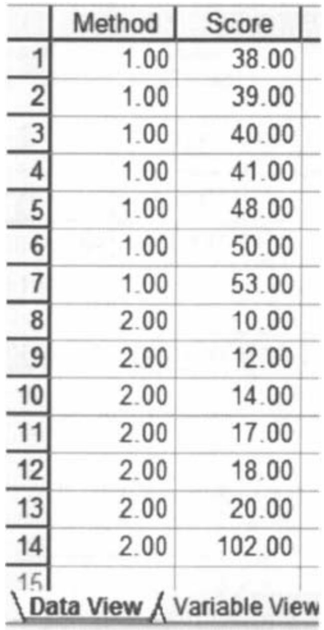


FIGURE 4.3

2. *Type in your values.*

   Click the "Data View" tab at the bottom of your screen as shown in Figure 4.4. Type in the values for both sets of data in the "Score" column. As you do so, type in the corresponding grouping variable in the "Method" column. For example, all of the values for "Method 2" are signified by a value of 2 in the grouping variable column that we called "Method".



	Method	Score
1	1.00	38.00
2	1.00	39.00
3	1.00	40.00
4	1.00	41.00
5	1.00	48.00
6	1.00	50.00
7	1.00	53.00
8	2.00	10.00
9	2.00	12.00
10	2.00	14.00
11	2.00	17.00
12	2.00	18.00
13	2.00	20.00
14	2.00	102.00
15		

FIGURE 4.4

3. *Analyze your data.*
   As shown in Figure 4.5, use the pull-down menus to choose "Analyze", "Nonparametric Tests", and "2 Independent Samples...".

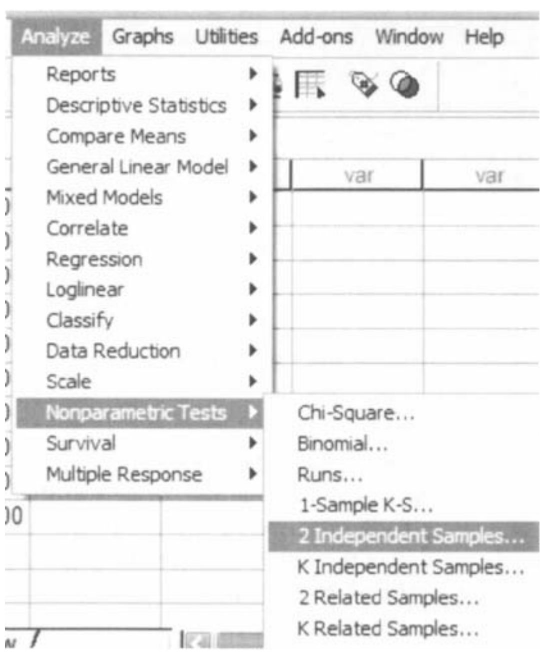


FIGURE 4.5

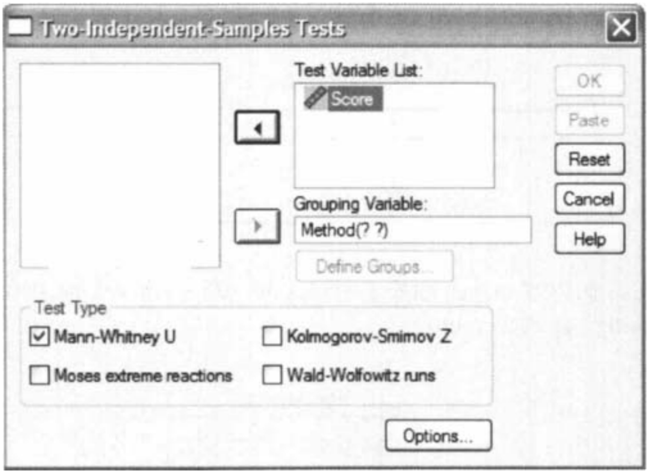


FIGURE 4.6

Use the top arrow button to place your variable with your data values, or dependent variable (DV), in the box labeled "Test Variable List:". Then, use the lower arrow button to place your grouping variable, or independent variable (IV), in the box labeled "Grouping Variable". As shown in Figure 4.6, we have placed the "Score" variable in the "Test Variable List" and the "Method" variable in the "Grouping Variable" box. Click on the "Define Groups..." button to assign a reference value to your independent variable (i.e., "Grouping Variable").

As shown in Figure 4.7, type 1 into the box next to "Group 1" and 2 in the box next to "Group 2". Then, click "Continue". This step references the value labels you created when you defined your grouping variable in step 1. Now that the groups have been assigned, click "OK" to perform the analysis.

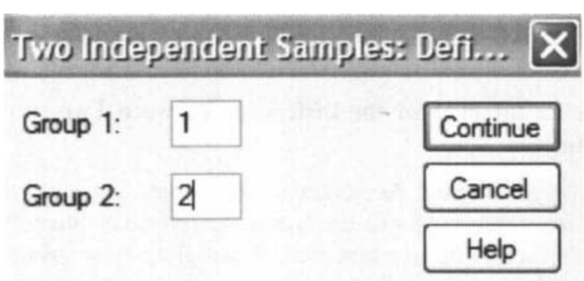


FIGURE 4.7

4. *Interpret the results from the SPSS Output window.*

Ranks

	Method	N	Mean Rank	Sum of Ranks
Score	Method 1	7	10.00	70.00
	Method 2	7	5.00	35.00
	Total	14		

The first SPSS output table provides the sum of ranks and sample sizes for comparing the two groups.

Test Statistics[b]

	Score
Mann-Whitney U	7.000
Wilcoxon W	35.000
Z	-2.236
Asymp. Sig. (2-tailed)	.025
Exact Sig. [2*(1-tailed Sig.)]	.026[a]

a. Not corrected for ties.

b. Grouping Variable: Method

The second SPSS output table provides the Mann–Whitney *U*-test statistic ($U = 7.0$). As described in Figure 4.2, it also returns a similar nonparametric statistic called the Wilcoxon *W*-test statistic ($W = 35.0$). Notice that the Wilcoxon *W* is the smaller of the two rank sums in the above table.

SPSS returns the critical *z*-score for large samples. In addition, SPSS calculates the two-tailed significance using two methods. The asymptotic significance is more appropriate with large samples. However, the exact significance is more appropriate with small samples or very poorly distributed data.

Based on the results from SPSS, the ranked reading comprehension test scores of the two methods were significantly different ($U = 7$, $n_1 = 7$, $n_2 = 7$, $p < 0.05$). The sum of ranks for Method 1 ($\sum R_1 = 70$) was larger than the sum of ranks for Method 2 ($\sum R_2 = 35$).

Note: Using the above procedure, SPSS always returns the two-tailed significance. If you were to need a one-tailed significance, simply divide by 2. The one-tailed exact significance in the above example is $0.026/2 = 0.013$.

4.3.3 Confidence Interval for the Difference Between Two Location Parameters

The American Psychological Association (2001) has suggested that researchers report the *confidence interval* for research data. A confidence interval is an inference to a population in terms of an estimation of sampling error. More specifically, it provides a range of values that fall within the population with a level of confidence of $100(1 - \alpha)\%$.

A median confidence interval can be constructed based on the difference between two independent samples. It consists of possible values of differences for which we do not reject the null hypothesis at a defined significance level of α.

The test depends on the following assumptions:

1. Data consist of two independent random samples: $X_1, X_2, \ldots, X_n$ from one population and $Y_1, Y_2, \ldots, Y_n$ from the second population.
2. The distribution functions of the two populations are identical except for possible location parameters.

To perform the analysis, set up a table that identifies all possible differences for each possible sample pair such that $D_{ij} = X_i - Y_j$ for (X_i, Y_j). Placing the values for X from smallest to largest across the top and the values for Y from smallest to largest down the side will eliminate the need to order the values of D_{ij} later.

The sample procedure presented below is based on the data from Table 4.2 (small data sample Mann–Whitney U-test) near the beginning of this chapter.

The values from Table 4.2 are arranged in Table 4.3 so that the Method 1 (X) scores are placed in order across the top and the Method 2 (Y) scores are placed in order down the side. Then, the $n_1 n_2$ differences are calculated by subtracting each Y value from each X value. The differences are shown in Table 4.3. Notice that the values of D_{ij} are ordered in the table from highest to lowest starting at the top right and ending at the bottom left.

TABLE 4.3

	X_i						
Y_j	38	39	40	41	48	50	53
10	28	29	30	31	38	40	43
12	26	27	28	29	36	38	41
14	24	25	26	27	34	36	39
17	21	22	23	24	31	33	36
18	20	21	22	23	30	32	35
20	18	19	20	21	28	30	33
102	−64	−63	−62	−61	−54	−52	−49

We use Table B.4 to find the lower limit of the confidence interval, L, and the upper limit, U. For a two-tailed test, L is the $w_{\alpha/2}$th smallest difference and U is the $w_{\alpha/2}$th largest difference that correspond to $\alpha/2$ for n_1 and n_2 for a confidence interval of $(1 - \alpha)$.

For our example, $n_1 = 7$ and $n_2 = 7$. For $0.05/2 = 0.025$, Table B.4 returns $w_{\alpha/2} = 9$. This means that the ninth values from the top and bottom mark the limits of the 95% confidence interval on both ends. Therefore, $L = 19$ and $U = 36$. Based on these results, we are 95% certain that the difference in population median is between 19 and 36.

4.3.4 Sample Mann–Whitney U-Test (Large Data Samples)

The previous comparison of teaching methods for reading recovery was repeated with fifth-grade students. The fifth grade used the same two methods. Method 1 was a pullout program in which the children were taken out of the classroom for 30 min. a day, 4 days a week. Method 2 was a small group program in which children were taught in groups of four or five for 45 min. a day in the classroom, 4 days a week. The students were tested using the same reading comprehension test after 4 weeks of the program. The test results are shown in Table 4.4.

TABLE 4.4

Method 1	Method 2
48	14
40	18
39	20
50	10
41	12
38	102
71	21
30	19
15	100
33	23
47	16
51	82
60	13
59	25
58	24
42	97
11	28
46	9
36	34
27	52
93	70
72	22
57	26
45	8
53	17

1. *State the null and research hypotheses.*

   The null hypothesis, shown below, states that there is no tendency of the ranks of one method to be systematically higher or lower than those of the other. The hypothesis is stated in terms of comparison of distributions, not means. The research hypothesis, shown below, states that the ranks of one method are systematically higher or lower than those of the other. Our research hypothesis is a two-tailed, nondirectional hypothesis because it indicates a difference, but in no particular direction.

The null hypothesis is

H_0: There is no tendency for ranks of one method to be significantly higher (or lower) than those of the other.

The research hypothesis is

H_A: The ranks of one method are systematically higher (or lower) than those of the other.

2. *Set the level of risk (or the level of significance) associated with the null hypothesis.*

   The level of risk, also called an alpha (α), is frequently set at 0.05. We will use an alpha of 0.05 in our example. In other words, there is a 95% chance that any observed statistical difference will be real and not due to chance.

3. *Choose the appropriate test statistic.*

   The data are obtained from two independent, or unrelated, samples of fifth-grade children being taught reading. Since we are comparing two unrelated, or independent, samples, we will use the Mann–Whitney *U*-test.

4. *Compute the test statistic.*

   First, combine and rank both data samples together (see Table 4.5). Next, compute the sum of ranks for each method. Method 1 is $\sum R_1$ and Method 2 is $\sum R_2$. Using Table 4.5,

$$\sum R_1 = 779$$

and

$$\sum R_2 = 496$$

Now, compute the U value for each sample. For sample 1,

$$\begin{aligned} U_1 &= n_1 n_2 + \frac{n_1(n_1+1)}{2} - \sum R_1 \\ &= 25(25) + \frac{25(25+1)}{2} - 779 = 625 + 325 - 779 \\ &= 171 \end{aligned}$$

and for sample 2,

$$\begin{aligned} U_2 &= n_1 n_2 + \frac{n_2(n_2+1)}{2} - \sum R_2 \\ &= 25(25) + \frac{25(25+1)}{2} - 496 = 625 + 325 - 496 \\ &= 454 \end{aligned}$$

The Mann–Whitney *U*-test statistic is the smaller of U_1 and U_2. Therefore, $U = 171$.

TABLE 4.5

Ordered Scores		
Rank	Score	Sample
1	8	Method 2
2	9	Method 2
3	10	Method 2
4	11	Method 1
5	12	Method 2
6	13	Method 2
7	14	Method 2
8	15	Method 1
9	16	Method 2
10	17	Method 2
11	18	Method 2
12	19	Method 2
13	20	Method 2
14	21	Method 2
15	22	Method 2
16	23	Method 2
17	24	Method 2
18	25	Method 2
19	26	Method 2
20	27	Method 1
21	28	Method 2
22	30	Method 1
23	33	Method 1
24	34	Method 2
25	36	Method 1
26	38	Method 1
27	39	Method 1
28	40	Method 1
29	41	Method 1
30	42	Method 1
31	45	Method 1
32	46	Method 1
33	47	Method 1
34	48	Method 1
35	50	Method 1
36	51	Method 1
37	52	Method 2
38	53	Method 1
39	57	Method 1
40	58	Method 1
41	59	Method 1
42	60	Method 1
43	70	Method 2
44	71	Method 1

TABLE 4.5 *(Continued)*

Ordered Scores		
Rank	Score	Sample
45	72	Method 1
46	82	Method 2
47	93	Method 1
48	97	Method 2
49	100	Method 2
50	102	Method 2

Since our sample sizes are large, we will approximate them to a normal distribution. Therefore, we will find a z-score for our data using a normal approximation. We must find the mean, $\bar{x}_U$, and the standard deviation, s_U, for the data.

$$\bar{x}_U = \frac{n_1 n_2}{2} = \frac{(25)(25)}{2}$$
$$= 312.5$$

and

$$s_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} = \sqrt{\frac{(25)(25)(25+25+1)}{12}} = \sqrt{\frac{31875}{12}}$$
$$= 51.54$$

Next, we use the mean, standard deviation, and the U-test statistic to calculate a z-score. Remember, we are testing the hypothesis that there is no difference in the ranks of the scores for two different methods of reading instruction for fifth-grade students.

$$z^* = \frac{U_i - \bar{x}_U}{s_U} = \frac{171 - 312.5}{51.54}$$
$$= -2.75$$

5. *Determine the value needed for rejection of the null hypothesis using the appropriate table of critical values for the particular statistic.*
   Table B.1 is used to establish the critical region of z-scores. For a two-tailed test with $\alpha = 0.05$, we must not reject the null hypothesis if $-1.96 \leq z^* \leq 1.96$.
6. *Compare the obtained value to the critical value.*
   We find that z^* is not within the critical region of the distribution, $-2.75 < -1.96$. Therefore, we reject the null hypothesis. This suggests a difference between Method 1 and Method 2.

7. *Interpret the results.*

   We rejected the null hypothesis, suggesting that a real difference exists between the two methods. In addition, since the sum of the ranks for Method 1 ($\sum R_1$) was larger than that for Method 2 ($\sum R_2$), we see that Method 1 had significantly higher scores.

   At this point, the analysis is limited to identifying the presence or absence of a significant difference between the groups. In other words, the statistical test's level of significance does not describe the strength of the treatment. The American Psychological Association (2001), however, has called for a measure of the strength called the *effect size.*

   We can consider the effect size for this large sample test to determine the degree of association between the groups. We can use Formula 4.5 to calculate the effect size. For the example, $z = -2.75$ and $n = 50$.

$$\begin{aligned} \text{ES} &= \frac{|z|}{\sqrt{n}} = \frac{|-2.75|}{\sqrt{50}} \\ &= 0.39 \end{aligned}$$

   Our effect size for the sample difference is 0.39. This value indicates a medium–high level of association between the teaching methods for the reading recovery program with fifth graders.

8. *Reporting the results.*

   For this example, two methods were used to provide fifth-grade students with reading instruction. Method 1 involved a pullout program and Method 2 involved a small group program. Using the ranked reading comprehension test scores, the results indicated a significant difference between the two methods ($U = 171$, $n_1 = 25$, $n_2 = 25$, $p < 0.05$). The sum of ranks for Method 1 ($\sum R_1 = 779$) was larger than the sum of ranks for Method 2 ($\sum R_2 = 496$). Moreover, the effect size for the sample difference was 0.39. Therefore, we can state that the data support the pullout program as a more effective reading program for teaching comprehension to fifth-grade children at this school.

4.4 EXAMPLES FROM THE LITERATURE

Below are varied examples of the nonparametric procedures described in this chapter. We have summarized each study's research problem and researchers' rationale(s) for choosing a nonparametric approach. We encourage you to obtain these studies if you are interested in their results.

- Odaci, H. (2007). Depression, submissive behaviors and negative automatic thoughts in obese Turkish adolescents. *Social Behavior & Personality: An International Journal, 35*(3), 409–416.

Odaci investigated depression, submissive social behaviors, and frequency of automatic negative thoughts in Turkish adolescents. Obese participants were compared with participants of normal weight. After the Shapiro–Wilk statistic revealed that the data were not normally distributed, Odaci applied a Mann–Whitney *U*-test to compare the groups.

- Bryant, B. K., & Trockel, J. F. (1976). Personal history of psychological stress related to locus of control orientation among college women. *Journal of Consulting and Clinical Psychology, 44*(2), 266–271.

  Bryant and Trockel investigated the impact of stressful life events on undergraduate females' locus of control. The authors compared accrued life changing units for participants with internal control against external using the Mann–Whitney *U*-test. This nonparametric procedure was selected since the data pertaining to stressful life events were ordinal in nature.

- Re, A. M., Pedron, M., & Cornoldi, C. (2007). Expressive writing difficulties in children described as exhibiting ADHD symptoms. *Journal of Learning Disabilities, 40*(3), 244–255.

  Re, Pedron, and Cornoldi investigated the expressive writing of children with attention-deficit/hyperactivity disorder (ADHD). The authors used a Mann–Whitney *U*-test to compare students showing symptoms of ADHD behaviors with a control group of students not displaying such behaviors. After examining their data with a Kolmogorov–Smirnov test, the researchers chose the nonparametric procedure due to significant deviations in the data distributions.

- Limb, G. E., & Organista, K. C. (2003). Comparisons between Caucasian students, students of color, and American Indian students on their views on social work's traditional mission, career motivations, and practice preferences. *Journal of Social Work Education, 39*(1), 91–109.

  In an effort to understand the factors that have motivated minority students to enter the social worker profession, Limb and Organista studied data from nearly 7000 students in California entering a social worker program. The authors used a Wilcoxon rank sum test to compare sums of student group ranks. They chose this nonparametric test due to a concern that statistical assumptions were violated regarding sample normality and homogeneity of variances.

- Schulze, E., & Tomal, A. (2006). The chilly classroom: Beyond gender. *College Teaching, 54*(3), 263–270.

  Schulze and Tomal examined classroom climate perceptions among undergraduate students. Since the student questionnaires used an interval scale, they analyzed their findings with a Mann–Whitney *U*-test.

- Hegedus, K. S. (1999). Providers' and consumers' perspective of nurses' caring behaviours. *Journal of Advanced Nursing, 30*(5), 1090–1096.

  Hegedus performed a pilot study to evaluate a scale designed to examine the caring behaviors of nurses. Care providers were compared with the consumers. She used a Wilcoxon rank sum test in her analysis because study participants were directed to rank the items on the scale.

4.5 SUMMARY

Two samples that are not related may be compared using a nonparametric procedure called the Mann–Whitney U-test (or the Wilcoxon rank sum test). The parametric equivalent to this test is known as the t-test for independent samples.

In this chapter, we described how to perform and interpret a Mann–Whitney U-test for both small samples and large samples. We also explained how to perform the procedure using SPSS. Finally, we offered varied examples of these nonparametric statistics from the literature. The next chapter will involve comparing more than two samples that are related.

4.6 PRACTICE QUESTIONS

1. The data in Table 4.6 were obtained from a reading-level test for first-grade children. Compare the performance gains of the two different methods for teaching reading.

TABLE 4.6

Method	Gain Score	Method	Gain Score
One-on-one	16	Small group	11
One-on-one	13	Small group	2
One-on-one	16	Small group	10
One-on-one	16	Small group	4
One-on-one	13	Small group	9
One-on-one	9	Small group	8
One-on-one	12	Small group	5
One-on-one	12	Small group	6
One-on-one	20	Small group	4
One-on-one	17	Small group	16

Use a two-tailed Mann–Whitney U-test with $\alpha = 0.05$ to determine which method was better for teaching reading. Report your findings.

2. A research study was conducted to see if an active involvement in a hobby had a positive effect on the health of a person who retires after age 65. The data in Table 4.7 describe the health (number of doctor visits in 1 year) for participants who are involved in a hobby almost daily and those who are not.

   Use a one-tailed Mann–Whitney U-test with $\alpha = 0.05$ to determine whether the hobby tends to reduce the need for doctor visits. Report your findings.

TABLE 4.7

No Hobby Group	Hobby Group
12	9
15	5
8	10
11	3
9	4
17	2
5	

3. Table 4.8 shows assessment scores of two different classes who are being taught computer skills using two different methods.

TABLE 4.8

Method 1	Method 2
53	91
41	18
17	14
45	21
44	23
12	99
49	16
50	10

Use a two-tailed Mann–Whitney U-test with $\alpha = 0.05$ to determine which method was better for teaching computer skills. Report your findings.

4. Two methods of teaching reading were compared. Method 1 used the computer to interact with the student and diagnose and remediate the student based upon misconceptions. Method 2 was taught using workbooks in classroom groups. Table 4.9 shows the data obtained on an assessment after 6 weeks of instruction. Calculate the effect size using the z-score from the analysis.
5. Two methods were used to provide instruction in science for the seventh grade. Method 1 included a lab each week and Method 2 had only classroom work with lecture and worksheets. Table 4.10 shows end-of-course test performance for the two methods. Construct a 95% median confidence interval based on the difference between two independent samples to compare the two methods.

TABLE 4.9

Method 1	Method 2
27	9
38	42
15	21
85	83
36	110
95	19
93	29
57	40
63	30
81	23
65	18
77	32
59	101
89	7
41	50
26	37
102	22
55	71
46	16
82	45
24	35
87	28
66	91
12	86
90	20

TABLE 4.10

Method 1	Method 2
15	8
23	15
9	10
12	13
18	17
22	5
17	18
20	7