6

COMPARING MORE THAN TWO UNRELATED SAMPLES: THE KRUSKAL–WALLIS *H*-TEST

6.1 OBJECTIVES

In this chapter, you will learn the following items.

- How to compute the Kruskal–Wallis *H*-test.
- How to perform contrasts to compare samples.
- How to perform the Kruskal–Wallis *H*-test and associated sample contrasts using SPSS.

6.2 INTRODUCTION

A professor asked her students to complete end-of-course evaluations for her psychology 101 class. She taught four sections of the course and wants to compare the evaluation results from each section. Since the evaluations were based upon a five-point rating scale, she decides to use a nonparametric procedure. Moreover, she recognizes that the four sets of evaluation results are independent, or unrelated. In other words, no single score in any single class is dependent upon any other score in any other class. This professor could compare her sections using the Kruskal–Wallis *H*-test.

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

The Kruskal–Wallis H-test is a nonparametric statistical procedure for comparing more than two samples that are independent, or not related. The parametric equivalent to this test is the one-way analysis of variance (ANOVA).

When the Kruskal–Wallis H-test leads to significant results, then at least one of the samples is different from the other samples. However, the test does not identify where the difference(s) occur. Moreover, it does not identify how many differences occur. To identify the particular differences between sample pairs, a researcher might use sample contrasts, or post hoc tests, to analyze the specific sample pairs for significant difference(s). The Mann–Whitney U-test is a useful method for performing sample contrasts between individual sample sets.

In this chapter, we will describe how to perform and interpret a Kruskal–Wallis H-test followed with sample contrasts. We will also explain how to perform the procedures using SPSS. Finally, we offer varied examples of these nonparametric statistics from the literature.

6.3 COMPUTING THE KRUSKAL–WALLIS H-TEST STATISTIC

The Kruskal–Wallis H-test is used to compare more than two independent samples. When stating our hypotheses, we state them in terms of the population. Moreover, we examine the population medians, θ_i, when performing the Kruskal–Wallis H-test.

To compute the Kruskal–Wallis H-test statistic, we begin by combining all of the samples and rank ordering the values together. Use Formula 6.1 to determine an H statistic.

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1) \tag{6.1}$$

where N is the number of values from all combined samples, R_i is the sum of the ranks from a particular sample, and n_i is the number of values from the corresponding rank sum.

The degrees of freedom, df, for the Kruskal–Wallis H-test are determined by using Formula 6.2.

$$\mathrm{df} = k - 1 \tag{6.2}$$

where df is the degrees of freedom and k is the number of groups.

Once the test statistic, H, is computed, it can be compared to a table of critical values (see Table B.6) to examine the groups for significant differences. However, if the number of groups, k, or the number of values in each sample, n_i, exceeds those available from the table, then a large sample approximation may be performed. Use a table with the chi-square distribution (see Table B.2) to obtain a critical value when performing a large sample approximation.

If ranking of values results in any ties, a ties correction is required. In that case, find a new H statistic by dividing the original H statistic by the ties correction. Use

Formula 6.3 to determine the ties correction value.

$$C_H = 1 - \frac{\sum(T^3 - T)}{N^3 - N} \tag{6.3}$$

where C_H is the ties correction, T is the number of values from a set of ties, and N is the number of values from all combined samples.

If the H statistic is not significant, then no differences exist between any of the samples. However, if the H statistic is significant, then a difference exists between at least two of the samples. Therefore, a researcher might use sample contrasts between individual sample pairs, or post hoc tests, to determine which of the sample pairs are significantly different.

When performing multiple sample contrasts, the Type I error rate tends to become inflated. Therefore, the initial level of risk, or α, must be adjusted. We recommend the Bonferroni procedure, shown in Formula 6.4, to adjust α.

$$\alpha_{\mathrm{B}} = \frac{\alpha}{k} \tag{6.4}$$

where α_{B} is the adjusted level of risk, α is the original level of risk, and k is the number of comparisons.

6.3.1 Sample Kruskal–Wallis H-Test (Small Data Samples)

Researchers were interested in studying the social interaction of different adults. They sought to determine if social interaction can be tied to self-confidence. The researchers classified 17 participants into three groups based on the social interaction exhibited. The participant groups were labeled

High: constant interaction; talks with many different people; initiates discussion.

Medium: interacts with a variety of people; some periods of isolation; tends to focus on fewer people.

Low: remains mostly isolated from others; speaks if spoken to, but leaves interaction quickly.

After the participants had been classified into the three social interaction groups, they were directed to complete a self-assessment of self-confidence on a 25-point scale. Table 6.1 shows the scores obtained by each of the participants, with 25 points being an indication of high self-confidence.

The original survey scores obtained were converted to an ordinal scale prior to the data analysis. Table 6.1 shows the ordinal values placed in the social interaction groups.

We want to determine if there is a difference between any of the three groups in Table 6.1. Since the data belong to an ordinal scale and the sample sizes are small ($n < 20$), we require a nonparametric test. The Kruskal–Wallis H-test is the best statistic to analyze the data and test the hypothesis.

TABLE 6.1

Original Ordinal Self-Confidence Scores Placed Within Social Interaction Groups

High	Medium	Low
21	19	7
23	5	8
18	10	15
12	11	3
19	9	6
20		4

1. *State the null and research hypotheses.*

   The null hypothesis, shown below, states that there is no tendency for self-confidence to rank systematically higher or lower for any of the levels of social interaction. The research hypothesis states that there is a tendency for self-confidence to rank systematically higher or lower for at least one level of social interaction compared to at least one of the other levels. We generally use the concept of "systematic differences" in the hypotheses.

   The null hypothesis is

   H_0: $\theta_L = \theta_M = \theta_H$

   The research hypothesis is

   H_A: There is a tendency for self-confidence to rank systematically higher or lower for at least one level of social interaction when compared to the other levels.

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 three independent, or unrelated, samples of adults who are being assigned to three different social interaction groups by observation. They are then being assessed using a self-confidence scale with a total of 25 points. The three samples are small with some violations of our assumptions of normality. Since we are comparing three independent samples, we will use the Kruskal–Wallis H-test.

4. *Compute the test statistic.*

   First, combine and rank the three samples together (see Table 6.2). Place the participant ranks in their social interaction groups to compute the sum of ranks, R_i, for each group (see Table 6.3). Next, compute the sum of ranks for each social interaction group. The ranks in each group are added to obtain a total R value for the group.

TABLE 6.2

Original Ordinal Score	Participant Rank	Social Interaction Group
3	1	Low
4	2	Low
5	3	Medium
6	4	Low
7	5	Low
8	6	Low
9	7	Medium
10	8	Medium
11	9	Medium
12	10	High
15	11	Low
18	12	High
19	13.5	Medium
19	13.5	High
20	15	High
21	16	High
23	17	High

For the High group,

$$R_{\mathrm{H}} = 10 + 12 + 13.5 + 15 + 16 + 17 = 83.5$$
$$n_{\mathrm{H}} = 6$$

For the Medium group,

$$R_{\mathrm{M}} = 3 + 7 + 8 + 9 + 13.5 = 40.5$$
$$n_{\mathrm{M}} = 5$$

For the Low group,

$$R_{\mathrm{L}} = 1 + 2 + 4 + 5 + 6 + 11 = 29$$
$$n_{\mathrm{L}} = 6$$

TABLE 6.3

Ordinal Data Ranks			
High	Medium	Low	
10	3	1	$N = 17$
12	7	2	
13.5	8	4	
15	9	5	
16	13.5	6	
17		11	

These R values are used to compute the Kruskal–Wallis H-test statistic (see Formula 6.1). The number of participants in each group is identified by a lowercase n. The total group size in the study is identified by the uppercase N.

Now, using the data from Table 6.3, compute the H-test statistic using Formula 6.1.

$$
\begin{aligned}
H &= \frac{12}{N(N+1)}\sum_{i=1}^{k}\frac{R_i^2}{n_i} - 3(N+1) \\
&= \frac{12}{17(17+1)}\left(\frac{83.5^2}{6}+\frac{40.5^2}{5}+\frac{29^2}{6}\right) - 3(17+1) \\
&= 0.0392\,(1162.04+328.05+140.17)-54 \\
&= 0.0392\,(1630.26)-54 \\
&= 63.93-54 \\
&= 9.93
\end{aligned}
$$

Since there was a tie involved in the ranking, correct the value of H. First, compute the ties correction (see Formula 6.2). Then, divide the original H statistic by the ties correction, C_H.

$$
\begin{aligned}
C_H &= 1-\frac{\sum(T^3-T)}{N^3-N} \\
&= 1-\frac{(2^3-2)}{17^3-17} \\
&= 1-\frac{(8-2)}{(4913-17)} \\
&= 1-0.0001 \\
&= 0.9988
\end{aligned}
$$

Next, we divide to find the corrected H statistic.

$$\text{corrected}\,H = \text{original}\,H/C_H = 9.93/0.9988 = 9.94$$

For this set of data, notice that the corrected H does not differ greatly from the original H. With the correction, $H = 9.94$.

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

   We will use the critical values table for the Kruskal–Wallis H-test (see Table B.6) since it includes the number of groups, k, and the number of samples, n, for our data. In this case, we look for the critical value for $k = 3$ and $n_1 = 6$, $n_2 = 6$, and $n_3 = 5$ with $\alpha = 0.05$. Table B.5 returns a critical value for the Kruskal–Wallis H-test of 5.76.

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

   The critical value for rejecting the null hypothesis is 5.76 and the obtained value is $H = 9.94$. If the critical value is less than or equal to the obtained value, we must reject the null hypothesis. If instead, the critical value exceeds the obtained value, we do not reject the null hypothesis. Since the critical value is less than the obtained value, we must reject the null hypothesis.

At this point, it is worth mentioning that larger samples often result in more ties. While comparatively small, as observed in step 4, corrections for ties can make a difference in the decision regarding the null hypothesis. If the H were near the critical value of 5.99 for df $= 2$ (e.g., $H = 5.80$), and the ties correction calculated to be 0.965, the decision would be to reject the null hypothesis with the correction ($H = 6.01$), but to not reject the null hypothesis without the correction. Therefore, it is important to perform ties corrections.

7. *Interpret the results.*

We rejected the null hypothesis, suggesting that a real difference in self-confidence exists between one or more of the three social interaction types. In particular, the data show that those who were classified as fitting the definition of the "Low" group were mostly people who reported poor self-confidence and those who were in the "High" group were mostly people who reported good self-confidence. However, describing specific differences in this manner is speculative. Therefore, we need a technique for statistically identifying difference between groups, or contrasts.

7a. *Sample contrasts, or post hoc tests.*

The Kruskal–Wallis H-test identifies if a statistical difference exists; however, it does not identify how many differences exist and which samples are different. To identify which samples are different and which are not, we can use a procedure called contrasts, or post hoc tests. An appropriate test to use when comparing two samples at a time is the Mann–Whitney U-test described in Chapter 3.

It is important to note that performing several Mann–Whitney U-tests has a tendency to inflate the Type I error rate. In our example, we would compare three groups, $k = 3$. At an $\alpha = 0.05$, the Type I error rate would equal $1 - (1 - 0.05)^3 = 0.14$.

To compensate for this error inflation, we suggest using the Bonferroni procedure (see Formula 6.4). With this technique, we use a corrected α with the Mann–Whitney U-tests to determine significant differences between samples. For our example,

$$\begin{aligned} \alpha_{\mathrm{B}} &= \frac{\alpha}{k} = \frac{0.05}{3} \\ &= 0.0167 \end{aligned}$$

When we compare the three samples with the Mann–Whitney U-tests and α_{B} (see Chapter 3), we obtain the following results presented in Table 6.4.

Using $\alpha_{\mathrm{B}} = 0.0167$, we notice that the High–Low group comparison is indeed significantly different. The Medium–Low group comparison is not significant. The High–Medium group comparison requires some judgment since it is difficult to tell if the difference is significant or not; the way the value is rounded could change the result.

TABLE 6.4

Group Comparison	Mann–Whitney U Statistic	Rank Sum Difference	Significance
High–Medium	2.5	$48.5 - 17.5 = 31.0$	0.017
Medium–Low	7.0	$38.0 - 28.0 = 10.0$	0.177
High–Low	1.0	$56.0 - 22.0 = 34.0$	0.004

Note: If you are not comparing all of the samples for the Kruskal–Wallis H-test, then k is only the number of comparisons you are making with the Mann–Whitney U-tests. Therefore, comparing fewer samples will increase the chances of finding a significant difference.

8. *Reporting the results.*

   The reporting of results for the Kruskal–Wallis H-test should include such information as sample size for all of the groups, the H statistic, degrees of freedom, and p-value's relation to α. For this example, three social interaction groups were compared: high ($n_H = 6$), medium ($n_M = 5$), and low ($n_L = 6$). The Kruskal–Wallis H-test was significant ($H_{(2)} = 9.94$, $p < 0.05$). To compare individual pairs of samples, contrasts may be used (see Chapter 3).

6.3.2 Performing the Kruskal–Wallis H-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 Friedman ANOVA described in Chapter 4, you cannot simply enter each sample into a separate column to execute the Kruskal–Wallis H-test. You must use a grouping variable. In Figure 6.1, the first variable is the grouping variable that we called "Group". The second variable that we called "Score" will have our actual values.

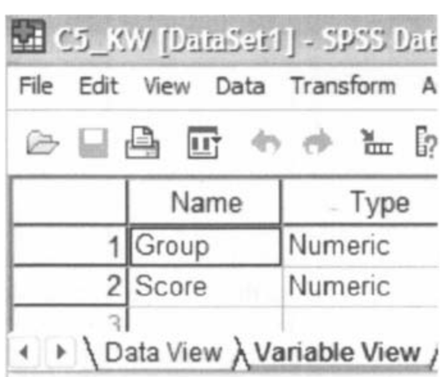

FIGURE 6.1

When establishing a grouping variable, it is often easiest to assign each group a whole number value. In our example, our groups are "High", "Medium", and "Low". Therefore, we must set our grouping variables for the variable "Group". First, we selected the "Values" column and clicked the gray square as shown in Figure 6.2. Then, we set a value of 1 to equal "High" and a value of 2 to equal "Medium". As soon as we click the "Add" button, we will have set "Low" equal to 3.

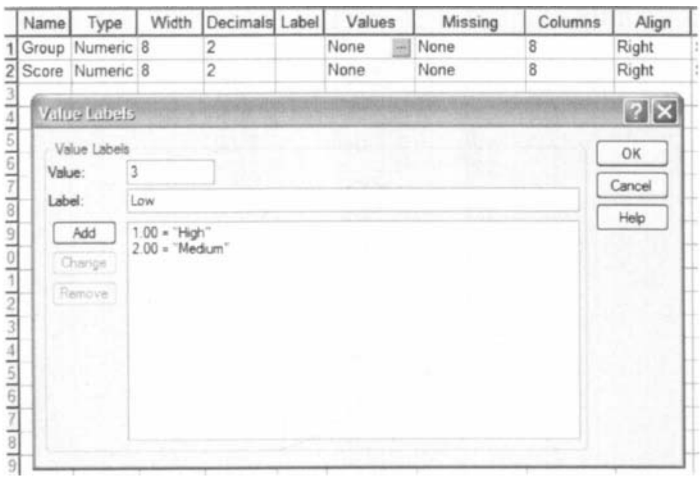

FIGURE 6.2

2. *Type in your values.*
   Click the "Data View" tab at the bottom of your screen as shown in Figure 6.3. Type in the values for all three samples in the "Score" column. As you do so, type in the corresponding grouping variable in the "Group" column. For example, all of the values for "High" are signified by a value of 1 in the grouping variable column that we called "Group".
3. *Analyze your data.*
   As shown in Figure 6.4, use the pull-down menus to choose "Analyze", "Nonparametric Tests", and "K Independent Samples...".
   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 6.5, we have placed the "Score" variable in the "Test Variable List" and the "Group" variable in the "Grouping Variable" box.
   Click on the "Define Range..." button to assign a reference value to your independent variable (i.e., "Grouping Variable").

	Group	Score
1	1.00	21.00
2	1.00	23.00
3	1.00	18.00
4	1.00	12.00
5	1.00	19.00
6	1.00	20.00
7	2.00	19.00
8	2.00	5.00
9	2.00	10.00
10	2.00	11.00
11	2.00	9.00
12	3.00	7.00
13	3.00	8.00
14	3.00	15.00
15	3.00	3.00
16	3.00	6.00
17	3.00	4.00

Data View | Variable View

FIGURE 6.3

FIGURE 6.4

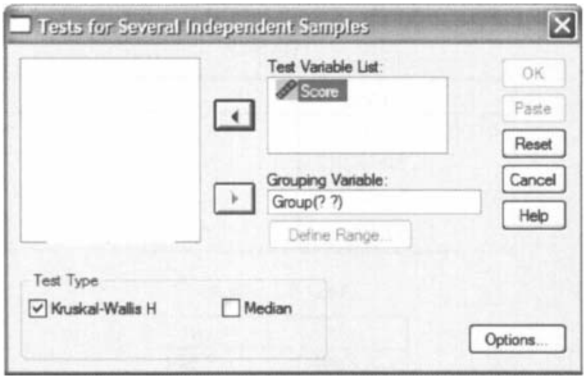

FIGURE 6.5

As shown in Figure 6.6, type 1 into the box next to "Minimum" and 3 in the box next to "Maximum". Then, click "Continue". This step references the value labels you defined when you established your grouping variable.

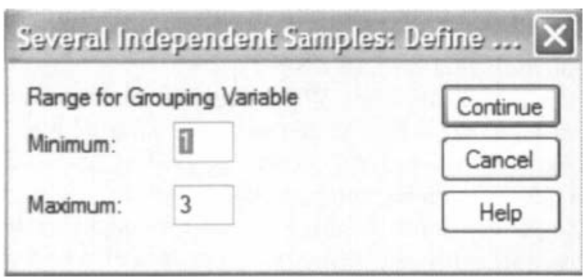

FIGURE 6.6

Now that the groups have been assigned (see Figure 6.7), click "OK" to perform the analysis.

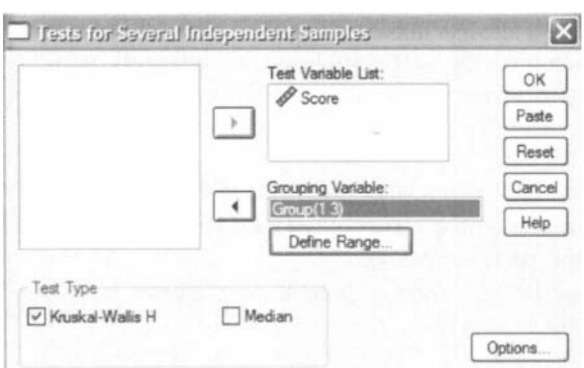

FIGURE 6.7

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

Ranks

	Group	N	Mean Rank
Score	High	6	13.92
	Medium	5	8.10
	Low	6	4.83
	Total	17	

Test Statistics[a,b]

	Score
Chi-Square	9.944
df	2
Asymp. Sig.	.007

a. Kruskal Wallis Test

b. Grouping Variable: Group

The first SPSS output table provides the mean ranks of groups and group sizes. The SPSS second output table provides the Kruskal–Wallis H-test statistic ($H = 9.944$). Since this test uses a chi-square distribution, SPSS calls the H statistic "Chi-Square". This table also returns the degrees of freedom ($\mathrm{df} = 2$) and the significance ($p = 0.007$).

Based on the results from SPSS, three social interaction groups were compared: high ($n_H = 6$), medium ($n_M = 5$), and low ($n_L = 6$). The Kruskal–Wallis H-test was significant ($H_{(2)} = 9.94$, $p < 0.05$). To compare individual pairs of samples, contrasts must be used.

Note: To perform Mann–Whitney U-tests for sample contrasts, simply use the grouping values you established when you defined your variables in step 1. Remember to use your corrected level of risk, α_B, when examining your significance.

6.3.3 Sample Kruskal–Wallis H-Test (Large Data Samples)

Researchers were interested in continuing their study of social interaction. In a new study, they examined the self-confidence of teenagers with respect to social interaction. Three levels of social interaction were based upon the following characteristics.

High: constant interaction; talks with many different people; initiates discussion.

Medium: interacts with a variety of people; some periods of isolation; tends to focus on fewer people.

Low: remains mostly isolated from others; speaks if spoken to, but leaves interaction quickly.

The researchers assigned each participant into one of the three social interaction groups. Researchers administered a self-assessment of self-confidence. The assessment instrument measured self-confidence on a 50-point ordinal scale. Table 6.5 shows the scores obtained by each of the participants, with 50 points indicating high self-confidence.

We want to determine if there is a difference between any of the three groups in Table 6.5. The Kruskal–Wallis H-test will be used to analyze the data.

TABLE 6.5

Original Self-Confidence Scores Placed Within Social Interaction Groups

High	Medium	Low
18	35	37
27	47	24
24	11	7
30	31	19
48	12	20
16	39	14
43	11	38
46	14	16
49	40	12
34	48	31
28	32	15
20	9	20
37	44	25
21	30	10
20	33	36
16	26	45
23	22	48
12	3	42
50	41	42
25	17	21
	8	
	10	
	41	

1. *State the null and research hypotheses.*

   The null hypothesis, shown below, states that there is no tendency for teen self-confidence to rank systematically higher or lower for any of the levels of social interaction. The research hypothesis states that there is a tendency for teen self-confidence to rank systematically higher or lower for at least one level of social interaction compared to at least one of the other levels. We generally use the concept of "systematic differences" in the hypotheses.

The null hypothesis is

H_0: $\theta_L = \theta_M = \theta_H$

The research hypothesis is

H_A: There is a tendency for teen self-confidence to rank systematically higher or lower for at least one level of social interaction when compared to the other levels.

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 three independent, or unrelated, samples of teenagers. They were assessed using an instrument with a 50-point ordinal scale. Since we are comparing three independent samples of values based on an ordinal scale instrument, we will use the Kruskal–Wallis H-test.
4. *Compute the test statistic.*
   First, combine and rank the three samples together (see Table 6.6).

TABLE 6.6

Original Ordinal	Participant Score Rank	Social Interaction Group
3	1	Medium
7	2	Low
8	3	Medium
9	4	Medium
10	5.5	Medium
10	5.5	Low
11	7.5	Medium
11	7.5	Medium
12	10	High
12	10	Medium
12	10	Low
14	12.5	Medium
14	12.5	Low
15	14	Low
16	16	High
16	16	High
16	16	Low
17	18	Medium
18	19	High
19	20	Low
20	22.5	High
20	22.5	High
20	22.5	Low

TABLE 6.6 (*Continued*)

Original Ordinal	Participant Score Rank	Social Interaction Group
20	22.5	Low
21	25.5	High
21	25.5	Low
22	27	Medium
23	28	High
24	29.5	High
24	29.5	Low
25	31.5	High
25	31.5	Low
26	33	Medium
27	34	High
28	35	High
30	36.5	High
30	36.5	Medium
31	38.5	Medium
31	38.5	Low
32	40	Medium
33	41	Medium
34	42	High
35	43	Medium
36	44	Low
37	45.5	High
37	45.5	Low
38	47	Low
39	48	Medium
40	49	Medium
41	50.5	Medium
41	50.5	Medium
42	52.5	Low
42	52.5	Low
43	54	High
44	55	Medium
45	56	Low
46	57	High
47	58	Medium
48	60	High
48	60	Medium
48	60	Low
49	62	High
50	63	High

TABLE 6.7

Ordinal Data Ranks		
High	Medium	Low
10	1	2
16	3	5.5
16	4	10
19	5.5	12.5
22.5	7.5	14
22.5	7.5	16
25.5	10	20
28	12.5	22.5
29.5	18	22.5
31.5	27	25.5
34	33	29.5
35	36.5	31.5
36.5	38.5	38.5
42	40	44
45.5	41	45.5
54	43	47
57	48	52.5
60	49	52.5
62	50.5	56
63	50.5	60
	55	
	58	
	60	

Place the participant ranks in their social interaction groups to compute the sum of ranks, R_i, for each group (see Table 6.7).

Next, compute the sum of ranks for each social interaction group. The ranks in each group are added to obtain a total R value for the group.

For the High group, $R_H = 709.5$ and $n_H = 20$.
For the Medium group, $R_M = 699$ and $n_M = 23$.
For the Low group, $R_L = 607.5$ and $n_L = 20$.

These R values are used to compute the Kruskal–Wallis H-test statistic (see Formula 6.1). The number of participants in each group is identified by a lowercase n. The total group size in the study is identified by the uppercase N. In this study, $N = 63$.

Now, using the data from Table 6.7, compute the H-test statistic using Formula 6.1.

$$
\begin{aligned}
H &= \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1) \\
&= \frac{12}{63(63+1)} \left(\frac{709.5^2}{20} + \frac{699^2}{23} + \frac{607.5^2}{20} \right) - 3(63+1) \\
&= 0.003\,(25169.51 + 21243.52 + 18452.81) - 192 \\
&= 0.003\,(64865.85) - 192 \\
&= 1.053
\end{aligned}
$$

Since there were ties involved in the ranking, correct the value of H. First, compute the ties correction (see Formula 6.2). There were 11 sets of ties with two values, 3 sets of ties with three values, and 1 set of ties with four values. Then, divide the original H statistic by the ties correction, C_H.

$$
\begin{aligned}
C_H &= 1 - \frac{\sum (T^3 - T)}{N^3 - N} \\
&= 1 - \frac{11(2^3-2) + 3(3^3-3) + (4^3-4)}{63^3 - 63} \\
&= 1 - \frac{189}{249984} \\
&= 1 - 0.0008 \\
&= 0.9992
\end{aligned}
$$

Next, we divide to find the corrected H statistic.

$$\text{corrected } H = \text{original } H / C_H = 1.053/0.9992$$

For this set of data, notice that the corrected H does not differ greatly from the original H. With the correction, $H = 1.054$.

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

   Since the data have at least one large sample, we will use the chi-square distribution (see Table B.2) to find the critical value for the Kruskal–Wallis H-test. In this case, we look for the critical value for df $= 2$ and $\alpha = 0.05$. Using the table, the critical value for rejecting the null hypothesis is 5.99.

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

   The critical value for rejecting the null hypothesis is 5.99 and the obtained value is $H = 1.054$. If the critical value is less than or equal to the obtained value, we must reject the null hypothesis. If instead, the critical value exceeds the obtained value, we do not reject the null hypothesis. Since the critical value exceeds the obtained value, we do not reject the null hypothesis.

7. *Interpret the results.*

   We did not reject the null hypothesis, suggesting that no real difference exists between any of the three groups. In particular, the data suggest that there is no difference in self-confidence between one or more of the three social interaction types.

8. *Reporting the results.*

   The reporting of results for the Kruskal–Wallis H-test should include such information as sample size for each of the groups, the H statistic, degrees of freedom, and p-value's relation to α. For this example, three social interaction groups were compared. The three social interaction groups were high ($n_H = 20$), medium ($n_M = 23$), and low ($n_L = 20$). The Kruskal–Wallis H-test was not significant ($H_{(2)} = 1.054$, $p > 0.05$).

6.4 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.

- Gömleksız, M. N., & Bulut, İ. (2007). An evaluation of the effectiveness of the new primary school mathematics curriculum in practice. *Educational Sciences: Theory & Practice, 7*(1), 81–94.

  Gömleksız and Bulut examined primary school teachers' views on the implementation and effectiveness of a new primary school mathematics curriculum. When they examined the data, some of the samples were found to be non-normal. For those samples, they used a Kruskal–Wallis H-test, followed by Mann–Whitney U-tests to compare unrelated samples.

- Finson, K. D., Pedersen, J., & Thomas, J. (2006). Comparing science teaching styles to students' perceptions of scientists. *School Science and Mathematics, 106*(1), 8–15.

  In Finson, Pedersen, and Thomas' study, the students of nine middle school teachers were asked to draw a scientist. Based on the drawings, students' perceptions of scientists were compared with their teachers' teaching styles using the Kruskal–Wallis H-test. Then, the samples were individually compared using the Mann–Whitney U-test. The researchers used nonparametric statistical analyses because only relatively small sample sizes of subjects were available.

- Belanger, N. D., & Desrochers, S. (2001). Can 6-month-old infants process causality in different types of causal events? *British Journal of Developmental Psychology, 19*(1), 11–21.

  Belanger and Desrochers investigated the nature of infants' ability to perceive event causality. The researchers noted that they chose nonparametric statistical tests because the data samples lacked a normal distribution based on results from a Shapiro–Wilk test. In addition, they stated that the sample sizes were small. A Kruskal–Wallis H-test revealed no significant differences between samples. Therefore, they did not perform any sample contrasts.

- Plata, M., & Trusty, J. (2005). Effect of socioeconomic status on general and at-risk high school boys' willingness to accept same-sex peers with LD. *Adolescence, 40*(157), 47–66.

  Plata and Trusty investigated high school boys' willingness to allow same-sex peers with learning disabilities (LD) to participate in school activities and out-of-school activities. The authors compared the willingness of 38 educationally successful and 33 educationally at-risk boys. The boys were from varying socioeconomic backgrounds. Due to the ordinal nature of data and small sample sizes among some samples, nonparametric statistics were used for the analysis. The Kruskal–Wallis H-test was chosen for multiple comparisons. When sample pairs were compared, the researchers performed a post hoc analysis of the differences between mean rank pairs using a multiple comparison technique.

6.5 SUMMARY

More than two samples that are not related may be compared using a nonparametric procedure called the Kruskal–Wallis H-test. The parametric equivalent to this test is known as the one-way analysis of variance. When the Kruskal–Wallis H-test produces significant results, it does not identify which or how many sample pairs are significantly different. The Mann–Whitney U-test, with a Bonferroni procedure to avoid Type I error rate inflation, is a useful method for comparing individual sample pairs.

In this chapter, we described how to perform and interpret a Kruskal–Wallis H-test followed with sample contrasts. We also explained how to perform the procedures using SPSS. Finally, we offered varied examples of these nonparametric statistics from the literature. The next chapter will involve comparing two variables.

6.6 PRACTICE QUESTIONS

1. A researcher conducted a study with $n = 15$ participants to investigate strength gains from exercise. The participants were divided into three groups and given one of three treatments. Participants' strength gains were measured and ranked. The rankings are presented in Table 6.8.

   Use a Kruskal–Wallis H-test with $\alpha = 0.05$ to determine if one or more of the groups are significantly different. If a significant difference exists, use two-tailed Mann–Whitney U-tests to identify which groups are significantly different. Use the Bonferroni procedure to limit the Type I error rate. Report your findings.

TABLE 6.8

Treatments		
I	II	III
7	13	12
2	1	5
4	7	16
11	8	9
15	3	14

2. A researcher investigated how physical attraction influences the perception among others of a person's effectiveness with difficult tasks. The photographs of 24 people were shown to a focus group. The group was asked to classify the photos into three groups: very attractive, average, and very unattractive. Then, the group ranked the photographs according to their impression of how capable they were of solving difficult problems. Table 6.9 shows the classification and rankings of the people in the photos (1: most effective; 24: least effective).

TABLE 6.9

Very Attractive	Average	Very Unattractive
1	3	11
2	4	15
5	8	16
6	9	18
7	13	20
10	14	21
12	19	23
17	22	24

Use a Kruskal–Wallis H-test with $\alpha = 0.05$ to determine if one or more of the groups are significantly different. If a significant difference exists, use two-tailed Mann–Whitney U-tests to identify which groups are significantly different. Use the Bonferroni procedure to limit the Type I error rate. Report your findings.

6.7 SOLUTIONS TO PRACTICE QUESTIONS

1. The results from the Kruskal–Wallis H-test are displayed in the SPSS Outputs below.

Ranks

	Treatment	N	Mean Rank
RankGain	Treatment 1	5	7.40
	Treatment 2	5	6.00
	Treatment 3	5	10.60
	Total	15	

Test Statistics[a,b]

	RankGain
Chi-Square	2.800
df	2
Asymp. Sig.	.247

a. Kruskal Wallis Test

b. Grouping Variable: Treatment

According to the data, the results from the Kruskal–Wallis H-test indicated that the three groups are not significantly different ($H_{(2)} = 2.800$, $p > 0.05$). Therefore, no follow-up contrasts are needed.

2. The results from the Kruskal–Wallis H-test are displayed in the SPSS Outputs below.

Ranks

	Classification	N	Mean Rank
Ranking	Very Attractive	8	7.50
	Average	8	11.50
	Very Unattractive	8	18.50
	Total	24	

Test Statistics[a,b]

	Ranking
Chi-Square	9.920
df	2
Asymp. Sig.	.007

a. Kruskal Wallis Test

b. Grouping Variable: Classification

According to the data, the results from the Kruskal–Wallis H-test indicated that one or more of the three groups are significantly different ($H_{(2)} = 9.920$, $p < 0.05$). Therefore, we must examine each set of samples with follow-up contrasts to find the differences between groups.

Based on the significance from the Kruskal–Wallis H-test, we compare the samples with Mann–Whitney U-tests. Since there are $k = 3$ groups, use $\alpha_B = 0.0167$ to avoid Type I error rate inflation. The results from the Mann–Whitney U-tests are displayed in the SPSS Outputs below

a. *Very attractive–attractive comparison*

Ranks

	Classification	N	Mean Rank	Sum of Ranks
Ranking	Very Attractive	8	7.00	56.00
	Average	8	10.00	80.00
	Total	16		

Test Statistics[b]

	Ranking
Mann-Whitney U	20.000
Wilcoxon W	56.000
Z	-1.260
Asymp. Sig. (2-tailed)	.208
Exact Sig. [2*(1-tailed Sig.)]	.234[a]

a. Not corrected for ties.

b. Grouping Variable: Classification

The results from the Mann–Whitney U-test ($U = 20.0$, $n_1 = 8$, $n_2 = 8$, $p > 0.0167$) indicated that the two samples were not significantly different.

b. *Attractive–very unattractive comparison*

Ranks

	Classification	N	Mean Rank	Sum of Ranks
Ranking	Average	8	6.00	48.00
	Very Unattractive	8	11.00	88.00
	Total	16		

Test Statistics[b]

	Ranking
Mann-Whitney U	12.000
Wilcoxon W	48.000
Z	-2.100
Asymp. Sig. (2-tailed)	.036
Exact Sig. [2*(1-tailed Sig.)]	.038[a]

a. Not corrected for ties.

b. Grouping Variable: Classification

The results from the Mann–Whitney U-test ($U = 12.0$, $n_1 = 8$, $n_2 = 8$, $p > 0.0167$) indicated that the two samples were not significantly different.

c. *Very attractive–very unattractive comparison*

Ranks

	Classification	N	Mean Rank	Sum of Ranks
Ranking	Very Attractive	8	5.00	40.00
	Very Unattractive	8	12.00	96.00
	Total	16		

Test Statistics[b]

	Ranking
Mann-Whitney U	4.000
Wilcoxon W	40.000
Z	-2.941
Asymp. Sig. (2-tailed)	.003
Exact Sig. [2*(1-tailed Sig.)]	.002[a]

a. Not corrected for ties.

b. Grouping Variable: Classification

The results from the Mann–Whitney U-test ($U = 4.0$, $n_1 = 8$, $n_2 = 8$, $p < 0.0167$) indicated that the two samples were significantly different.