COMPARING MORE THAN TWOUNRELATED SAMPLES:

THE KRUSKAL–WALLIS H-TEST

OBJECTIVE

In this lecture, you will learn the following items:

- How to compute the Kruskal–Wallis H-test.
- • How to perform contrasts to compare samples

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 on 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 on any other score in any other class. This professor could compare her sections using the Kruskal–Wallis *H*-test.

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) occurs. Moreover, it does not identify how many differences occur. In order 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 per- forming sample contrasts between individual sample sets.

In this lecture, we will describe how to perform and interpret a Kruskal–Wallis H-test followed with sample contrasts.

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 1 to determine an H statistic:

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

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 2:

$$df = k - 1 \qquad (2)$$

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 with a table of critical values (Table B.6) to examine the groups for significant differences.

However, if the number of groups, k, or the numbers of values in each sample, n_i, exceed those available from the table, then a large sample approximation may be performed. Use a table with the $\chi 2$ distribution (Table B.2) to obtain a critical value when performing a large sample approximation.

If ranking of values results in any ties, a tie correction is required. In that case, find a new H statistic by dividing the original H statistic by the tie correction. Use Formula 3 to determine the tie correction value;

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

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 , must be adjusted. We demonstrate the Bonferroni procedure, shown in Formula 4, to adjust :

$$\alpha_B = \frac{\alpha}{k} \tag{4}$$

where α_B is the adjusted level of risk, α is the original level of risk, and k is the number of comparisons.

Example

Kruskal–Wallis *H*-Test

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 as follows:

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 1 shows the scores obtained by each of the participants, with 25 points being an indication of high self-confidence.

TABLE 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

The original survey scores obtained were converted to an ordinal scale prior to the data analysis.

Table 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 1.

Since the data belong to an ordinal scale and the sample sizes are small ($n < 20$), we will use a nonparametric test.

The Kruskal–Wallis H-test is a good choice to analyze the data and test the hypothesis.

1. State the Null and Research Hypotheses

The null hypothesis 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 than at least one of the other levels.

We generally use the concept of "systematic differences" in the hypotheses.

The null hypothesis is

$$H_O: \theta_L = \theta_M = \theta_H$$

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 with the other levels.

2. Set the Level of Risk (or the Level of Signiicance) Associated with the Null Hypothesis

The level of risk, also called an alpha (α), is frequently set at 0.05. We will use $\alpha = 0.05$ 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 2).

TABLE 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

Place the participant ranks in their social interaction groups to compute the sum of ranks R_i for each group (see Table 3).

TABLE 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	

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,

$$R_H = 10 + 12 + 13.5 + 15 + 16 + 17 = 83.5$$
$$n_H = 6$$

For the medium group,

$$R_M = 3 + 7 + 8 + 9 + 13.5 = 40.5$$
$$n_M = 5$$

For the low group,

$$R_L = 1 + 2 + 4 + 5 + 6 + 11 = 29$$
$$n_L = 6$$

These R-values are used to compute the Kruskal–Wallis H-test statistic (see Formula 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 3, compute the H-test statistic using Formula 1:

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

$$H = \frac{12}{17(17+1)} \left(\frac{83.5^2}{6} + \frac{40.5^2}{5} + \frac{29^2}{6} \right) - 3(17+1)$$

$$= 0.0392(1162.04 + 328.05 + 140.17) - 54 = 0.0392(1630.26) - 54 = 63.93 - 54$$

$$H = 9.93$$

Since there was a tie involved in the ranking, correct the value of H. First, compute the tie correction (see Formula 2).

Then, divide the original H statistic by the ties correction C_H:

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

$$C_H = 0.9988$$

Next, we divide to find the corrected H statistic:

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

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 value table for the Kruskal–Wallis H-test (Table B.6) since it includes the number of groups, k, and the numbers 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 = 0.05.

Table B.5 returns a critical value for the Kruskal–Wallis H-test of 5.76.

6. Compare the Obtained Value with 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 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 tie 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 tie 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.

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.

Methods for comparing two samples at a time are described in before.

The examples in this lecture compare unrelated samples so we will use the Mann–Whitney U-test.

It is important to note that performing several two-sample tests has a tendency to inflate the type I error rate. In our example, we would compare three groups, $k = 3$. At $\alpha = 0.05$, the type I error rate would be $1 - (1 - 0.05)^3 = 0.14$.

To compensate for this error inflation, we demonstrate the Bonferroni procedure (Formula 4).

With this technique, we use a corrected with the Mann–Whitney U-tests to determine significant differences between samples.

For our example,

$$\alpha_B = \frac{\alpha}{k} = \frac{0.05}{3}$$
$$\alpha_B = 0.0167$$

When we compare each set of samples with the Mann–Whitney U-tests and use B, we obtain the following results presented in Table 4.

TABLE 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

Since $\alpha_B = 0.0167$, 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 off could change the result.

Note that 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 .

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 ($H(2) = 9.94$, $p < 0.05$).

In order to compare each set of samples, contrasts may be used as described earlier in this lecture.

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 (ANOVA).

When the Kruskal–Wallis H-test produces significant results, it does not identify which nor 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 lecture, we described how to perform and interpret a Kruskal–Wallis H-test followed with sample contrasts.