Kruskal-Wallis Test

Omnibus Test

The **Kruskal-Wallis H test** is a non-parametric test which is used in place of a one-way ANOVA. Essentially it is an extension of the Wilcoxon Rank-Sum test to more than two independent samples.

Although, as explained in Assumptions for ANOVA, one-way ANOVA is usually quite robust, there are many situations where the assumptions are sufficiently violated and so the Kruskal-Wallis test becomes quite useful: in particular, when:

- Group samples strongly deviate from normal; this is especially relevant when sample sizes are small and unequal and data are not symmetric.
- Group variances are quite different because of the presence of outliers

If the assumptions of ANOVA are satisfied, then the Kruskal-Wallis test is less powerful than ANOVA, and so you should use ANOVA. This is also the case when a transformation can be used to meet the ANOVA assumptions. When the homogeneity assumption fails, Welch's ANOVA is often preferred over the Kruskal-Wallis test.

Some characteristics of Kruskal-Wallis test are:

- The assumptions are similar to those for the Mann-Whitney test: independent group samples, data in each group is randomly selected and data is at least ordinal
- No assumptions are made about the type of underlying distribution, although see below
- Each group sample has at least 5 elements.
- No population parameters are estimated, and so there are no confidence intervals.

The Kruskal-Wallis tests is actually testing the null hypothesis that the populations from which the group samples are selected are equal in the sense that none of the group populations is **dominant** over any of the others. A group is dominant over the others if when one element is draw at random from each of the group populations, it is more likely that the largest element is in that group.

> H_0: the group populations have equal dominance; i.e. when one element is drawn at random from each group population, the largest (or smallest, or second smallest, etc.) element is equally likely to come from any one of the group populations

> H_1: At least one of the group populations is dominant over the others

When the group samples have the same shape (and so presumably this is reflective of the corresponding population distributions), then the null hypothesis can be viewed as a statement about the group medians.

> H_0: the group population medians are equal

> H_1: the group population medians are not equal

An indication that the population distributions have the same shape (except that possibly there is a shift to the right or left among them) is that the box plots are similar, except that the box and whiskers among them may be at different heights. Another indication is that the group histograms or QQ plots look similar (although not necessarily indicating normality).

Property 1: Define the test statistic

$$H = \frac{12}{n(n+1)} \sum\nolimits_{j=1}^{k} \frac{R_j^2}{n_j} - 3(n+1)$$

where k = the number of groups, n_j is the size of the jth group, R_j is the rank sum for the jth group and n is the total sample size, i.e.

$$n = \sum\nolimits_{j=1}^{k} n_j$$

Then

$$H \sim \chi^2(k-1)$$

provided $n_j \geq 5$ based on the following null hypothesis:

> H_o: The distribution of scores is equal across all groups

Observation: If the assumptions of ANOVA are satisfied, then the Kruskal-Wallis test is less powerful than ANOVA.

An alternative expression for H is given by

$$H = \frac{12}{n(n+1)} SS'_B$$

where SS'_B is the sum of squares between groups using the ranks instead of raw data. This is based on the fact that $\frac{12(k-1}{n(n+1)}$ is the expected value (i.e. mean) of the distribution of SS'_B.

If there are small sample sizes and many ties, a corrected Kruskal-Wallis test statistic $H' = H/T$ gives better results where

$$T = 1 - \frac{1}{n^3 - n} \sum (f^3 - f)$$

Here the sum is taken over all scores where ties exist and f is the number of ties at that level.

Example 1: A cosmetic company created a small trial of a new cream for treating skin blemishes. It measured the effectiveness of the new cream compared to the leading cream on the market and a placebo. Thirty people were put into three groups of 10 at random, although just before the trial began 2 people from the control group and 1 person from the test group for the existing cream dropped out. The left side of Figure 1 shows the number of blemishes removed from each person during the trial.

	A	B	C	D	E	F	G	H	I
2						Shapiro-Wilk Test			
3		New	Old	Control					
4		81	48	18			*New*	*Old*	*Control*
5		32	31	49		W-stat	0.839268	0.964031	0.811959
6		42	25	33		p-value	0.043242	0.839419	0.038381
7		62	22	19		alpha	0.05	0.05	0.05
8		37	30	24		normal	no	yes	no
9		44	30	17					
10		38	32	48		d'Agostino-Pearson			
11		47	15	22					
12		49	40			DA-stat	10.49939	0.476894	1.954107
13		41				p-value	0.005249	0.787851	0.376419
14	mean	47.3	30.333	28.75		alpha	0.05	0.05	0.05
15	var	206.68	92.75	173.64		normal	no	yes	yes

Figure 1 – Blemish treatment data

Based on the Shapiro-Wilk test, shown on the right side of the figure, we see that two of the groups are not normally distributed. This conclusion is confirmed from the QQ plots (not shown here). We, therefore decide to use the Kruskal-Wallis test instead of ANOVA.

From the box plots shown in Figure 2, we observe, that although the group distributions don't have the exact same shape (consistent with the fact that two are not normally distributed, while one is normally distributed), their shapes are fairly similar (although the values for the New group are larger than for the other two groups). Thus, we can use Kruskal-Wallis to test the null hypothesis that none of the groups is dominant over the others, and perhaps even that the group medians are equal.

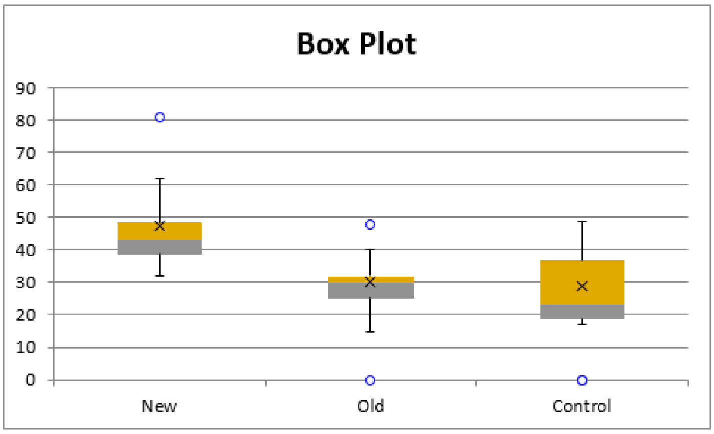

Figure 2 – Box plot comparisons

We now carry out the Kruskal-Wallis test as shown in Figure 3. Using the RANK.AVG function (or the RANK_AVG function for Excel 2007 users), we obtain the ranks of each of the raw scores, as shown in range G4:I13. E.g. cell I4 contains the formula =IF(ISNUMBER(D4), RANK.AVG(D4,$B$4:$D$13,1),"").

We next calculate the sum of the ranks for each group, namely R_1 = 199, R_2 = 96.5 and R_3 = 82.5. Next we square each of these values and divide by the number of elements in the corresponding group to obtain the figures shown in range G16:I16. The remaining formulas in the figure are shown in column L (corresponding to formulas in column J).

	F	G	H	I	J	K	L
3		New	Old	Control			
4		27	22.5	3			
5		12.5	11	24.5			
6		19	8	14			
7		26	5.5	4			
8		15	9.5	7			
9		20	9.5	2			
10		16	12.5	22.5			
11		21	1	5.5			
12		24.5	17				
13		18					
14	Rank Sums R	199	96.5	82.5			
15	Group Size n	10	9	8	27		=SUM(G15:I15)
16	R^2/n	3960.1	1034.7	850.78	5845.58		=SUM(G16:I16)
17	H				8.78692		=12*J16/(J15*(J15+1))-3*(J15+1)
18	df				2		=COUNTA(G3:I3)-1
19	p				0.01236		=CHISQ.DIST.RT(J17,J18)
20	α				0.05		
21	sig				yes		=IF(J19<J20,"yes","no")

Figure 3 – Kruskal-Wallis test

Since p-value = .01236 < .05 = α, we reject the null hypothesis, and conclude there is significant difference between the three cosmetics.

Note that we can perform a one-way ANOVA on the ranks (i.e. the data in range G3:I13) using Excel's **ANOVA: One Factor** data analysis tool (or the Real Statistic data analysis tool) to find SS_B. This provides an alternative way of calculating (see Figure 4) since H is equal to

$$H = \frac{SS_B}{n(n+1)/12}$$

	N	O	P	Q	R	S	T
3	Anova: Single Factor						
4							
5	SUMMARY						
6	*Groups*	*Count*	*Sum*	*Average*	*Variance*		
7	New	10	199	19.9	23.15556		
8	Old	9	96.5	10.72222	39.19444		
9	Control	8	82.5	10.3125	79.99554		
10							
11							
12	ANOVA						
13	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
14	Between Groups	553.5757	2	276.7878	6.139901	0.007024	3.402826
15	Within Groups	1081.924	24	45.08018			
16							
17	Total	1635.5	26				
18							
19	SS_B	553.5757					
20	n(n+1)/12	63					
21	H	8.786916					

Figure 4 – ANOVA on ranked data

Real Statistics Functions: The Real Statistics Resource Pack contains the following functions:

KRUSKAL(R1, *ties*) = value of H on the data (without headings) contained in range R1 (organized by columns).

KTEST(R1, *ties*) = p-value of the Kruskal-Wallis test on the data (without headings) contained in range R1 (organized by columns).

When *ties* = TRUE (default) then a ties correction is applied.

For Example 1, KRUSKAL(B5:D14) = 8.7869 and KTEST(B5:D14) = .01236.

The resource pack also provides the following array function:

KW_TEST(R1, *lab, ties*) = the 4 × 1 range consisting of the values for *H, H′, df*, p-value if *lab* = FALSE (default). If *lab* = TRUE then an extra column is added containing labels. If *ties* = TRUE (default) then a ties correction is applied (thus *H′* = *H* if no ties correction is applied).

Real Statistics Data Analysis Tool: The Real Statistics Resource Pack provides a data analysis tool to perform the Kruskal-Wallis test.

To use the tool for Example 1, press **Ctrl-m** and double click on **Analysis of Variance** (or click on the **Anova** tab if using the Multipage interface) and select **Single Factor Anova**. When a dialog box similar to that shown in Figure 1 of ANOVA Analysis Tool appears, enter B3:D13 in the **Input Range**, check **Column headings included with data**, select the **Kruskal-Wallis** option and click on **OK**.

The output is shown in Figure 5.

	V	W	X	Y	Z
5	Kruskal-Wallis Test				
6					
7		New	Old	Control	
8	median	43	30	23	
9	rank sum	199	96.5	82.5	
10	count	10	9	8	27
11	r^2/n	3960.1	1034.694	850.7813	5845.576
12	H				8.786916
13	H-ties				8.800347
14	df				2
15	p-value				0.012275
16	alpha				0.05
17	sig				yes

Figure 5 – Kruskal-Wallis data analysis

The *H′* value (including a ties correction) can be calculated by =KRUSKAL(B4:D13) and the corresponding p-value by =KTEST(B4:D13). In fact, the range Z12:Z15 can be calculated by =KW_TEST(B4:D13).

Follow-up Tests

If the Kruskal-Wallis Test shows a significant difference between the groups, then pairwise comparisons or contrasts can be used to pinpoint the difference(s) as described following a single factor ANOVA. It is important to reduce familywise Type I error.

For more information about these follow-up tests and how to perform them in Excel, click on any of the following links:

- Nemenyi Test
- Dunn's Test
- Schaich-Hamerle Test

- [Conover Test]
- [Steel Test]
- [Pairwise Mann-Whitney tests]
- [Contrasts]

136 Responses to *Kruskal-Wallis Test*

Dima Alnajar *says:*
January 27, 2018 at 9:51 am

Hallo dear charles,
Thanks for explaining Kruskal wallis test very nicely. I am a PhD student in plant pathology and i am studying the resistance of different plant varieties to a fungal disease. I monitor the lesions on the plant´s stem (length of leasion L, girdling G and profundity P) each in classes from 1-9 then i but them in an equation to find the volume of diseased tissue (HR=1-P/9 and then
Volume of Diseased Tissue VDT=(1-HR^2)*G/9*L.
i had two reasons to apply kruskal wallis test. First, I supposed that my results of VDT are still categorial after applying the equation. Second, the normal distribution was violated in many varieties so i could not apply ANOVE. My basic Problem is that i can not explain what does it mean to have significant results in the concrete situation in my experiment (Probably it means that the means of ranks are from different distribution) but still in my expirement i can not put it in words. Also in my presentation, how can i present my results in Graphs? i am using XLSTAT, should i present the means of ranks and put the letters that show significant differences above them? or can i use the means of the value. My supervisor told me that he is interested to see a graph with the means of VDT and not of the ranks. and now i do not know how to present a correct graph.
Thanks alot in advance for your help.
Best regards,
Dima

Reply

Charles *says:*
January 27, 2018 at 10:38 am

Dima,
1. The null hypothesis being tested by the KW test is: The distribution of scores is equal across all groups. This means that the group samples are coming from populations with the same distribution (including their parameters).
You don't need to mention ranks when stating the hypothesis.
2. The graphs should be of the data and not the ranks.
Charles

Reply