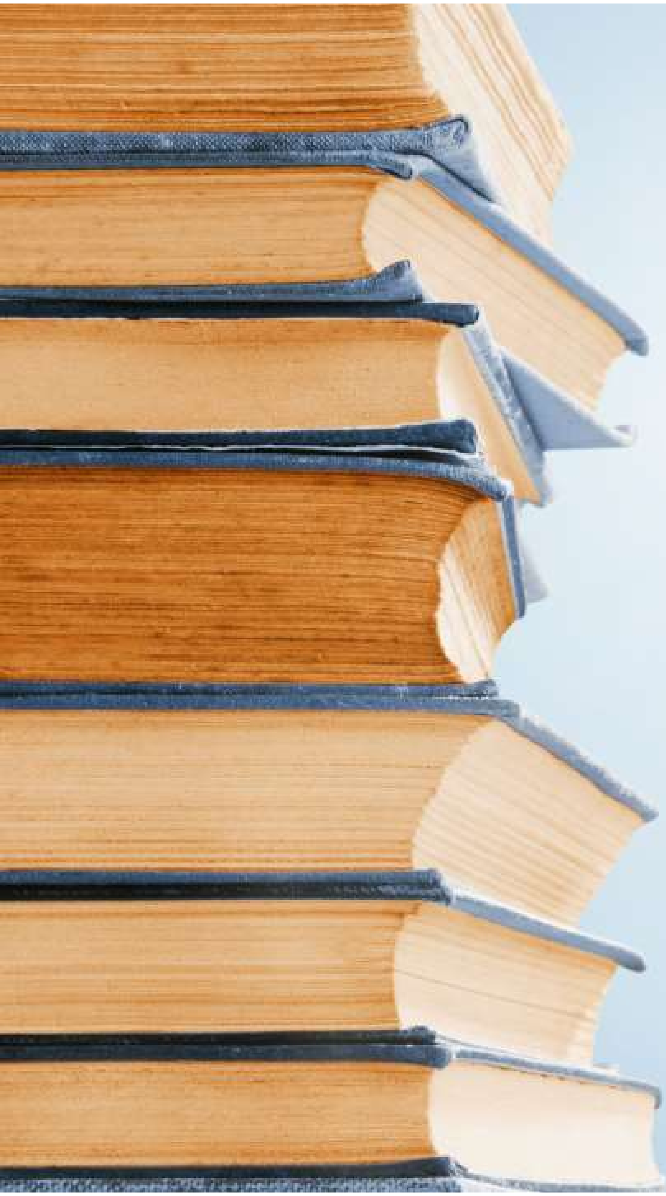

Kruskal-Wallis Test

Dr.E.S.Jeevanand,
Associate Professor
Union Christian College. Aluva

Kruskal-Wallis Test

The Kruskal-Wallis Test was developed by Kruskal and Wallis (1952) jointly and is named after them. The Kruskal-Wallis test is a nonparametric (distribution free) test, and is used when the assumptions of one way ANOVA are not met. They both assess for significant differences on a continuous dependent variable by a grouping independent variable (with three or more groups). In the ANOVA, we assume that distribution of each group is normally distributed and there is approximately equal variance on the scores for each group. However, in the Kruskal-Wallis Test, we do not have any of these assumptions.

One way ANOVA is a statistical data analysis technique that is used to test the equality of the mean of three or more independent variable

Kruskal-Wallis Test

- The **Kruskal-Wallis** H **test** (sometimes also called the "one-way ANOVA on ranks") is a rank-based nonparametric **test** that can be **used to** determine if there are statistically significant differences between two or more groups of an independent variable on a continuous or ordinal dependent variable. The Kruskal-Wallis one-way ANOVA is a non-parametric method for comparing k independent samples. It is roughly equivalent to a parametric one way ANOVA with the data replaced by their ranks. Since ranking is conditional upon your observed values, so is this test. The null and alternative hypothesis in this case are
- **Null hypothesis:** The samples are from identical populations.
- **Alternative hypothesis:** The sample comes from different populations.

Kruskal-Wallis Test

The Kruskal-Wallis is also used as a test of equality of medians or even means. In the latter case, in addition to the distributional assumptions mentioned above, observations are also assumed to be distributed symmetrically.

- The null and alternative hypothesis in this case are
- **Null hypothesis:** Null hypothesis assumes that the k samples median are equal. (i.e. H_0: $M_1 = M_2 = \cdots = M_k$)
- **Alternative hypothesis:** Alternative hypothesis assumes that some among the k samples median are different.

The outcome of the **Kruskal–Wallis test** tells you if there are differences among the medians of some of the k groups, but doesn't tell you which groups are different from other groups. In order to determine which groups are different from others, Mann Whitney U test can be conducted.

Kruskal-Wallis Test- Assumptions

- The assumptions of the Kruskal-Wallis test are similar to those for the Mann-Whitney test.
- Samples are random samples, or allocation to treatment group is random.
- The two samples are mutually independent. (independence within each sample and mutual independence among samples)
- The measurement scale is at least ordinal, and the variable is continuous.
- If the test is used as a test of dominance, it has no distributional assumptions. If it used to compare medians, the distributions must be similar apart from their locations.

Kruskal-Wallis Test
Procedure for carrying out the test

- Combine the observations in the k samples into a single pooled 'null' sample, retaining the information on the source of each observation.
- Under each observation, write down X or Y or Z etc. (or some other relevant symbol) to indicate which sample they are from.
- Assign ranks to the pooled sample after arrange the data in ascending order. If two values are the same, they both get the average of the two ranks for which they tie - in other words use mean ranks for tied observations, not sequential ranks. (i.e. the ranking is same as that of Mann-Whitney test)
- Once this is complete, ranks or the different samples are separated and summed up as R_1 R_2 R_3, etc.

Kruskal-Wallis Test
Procedure for carrying out the test

- Compute the Kruskal-Wallis test statistic (*H*).

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

- Where, $n = \sum_{i=1}^{k} n_i$ Type equation here. total number of observations in all samples Kruskal-Wallis Test statistic is approximately a chi-square distribution, with *k*-1 degree of freedom where n_i should be greater than 5.

- Decision criteria is

Reject H_0 if $H \geq \chi_\alpha^2$

- Where for a given α, χ_α^2 is the tabled value obtained from the chi-square table, with *k*-1 degree of freedom.

Example I

- Three products received the following performance rating by a panel of 20 customers. Use the Kurskal-Wallis test to determine whether there is significant difference in the performance ratings for the product.
- Here the hypothesis are
- H_0: the performance ratings are same for the product.
- H_1: the performance ratings are different for the product.

Product		
A	B	C
25	60	50
70	20	70
60	30	60
85	15	80
95	40	90
90	35	70
80		75

Example I ranking for pooled data

Rating	Product	Rank	Rating	Product	Rank
15	B	1	70	A	12
20	B	2	70	C	12
25	A	3	70	C	12
30	B	4	75	C	14
35	B	5	80	A	15.5
40	B	6	80	C	15.5
50	C	7	85	A	17
60	A	9	90	A	18.5
60	B	9	90	C	18.5
60	C	9	95	A	20

Example I

Product A	Rank	Product B	Rank	Product C	Rank
25	3	60	9	50	7
70	12	20	2	70	12
60	9	30	4	60	9
85	17	15	1	80	15.5
95	20	40	6	90	18.5
90	18.5	35	5	70	12
80	15.5			75	14
Total	R_1=95		R_2=27		R_3=88

Example I

Here $n_1 = 7$, $n_2 = 6$, $n_3 = 7$, $n = \sum_{i=1}^{k} n_i = 7+6+7 = 20$, $k=3$

$R_1 = 95$, $R_2 = 27$ and $R_3 = 88$

$$\sum_{i=1}^{k} \frac{R_i^2}{n_i} = \frac{95^2}{7} + \frac{27^2}{6} + \frac{88^2}{8} = 2517.071$$

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

$$= \frac{12}{20(20+1)} \times 2507.071 - 3(20+1) = 8.916.$$

The tabled value χ_α^2 for $\alpha=0.05$ and k=3-1 degrees of freedom is 5.99. Since the calculated value $H > \chi_\alpha^2$ test is significant. So we reject H_0 and conclude that the performance ratings are different for the product.

Example II

The marks of statistics for selected 5 students from three colleges are given in the table. Test whether the performance of the students of the three college are same or not.

- Here the hypothesis are
- H_0: $M_1 = M_2 = M_3$
- H_1: some of the colleges differs in median marks.

Where M_i is the median mark of i^{th} college.

College		
A	B	C
50	80	60
62	95	45
75	98	30
48	87	58
65	90	57

Example II ranking for pooled data

Marks	College	Rank	Marks	College	Rank
30	C	1	65	A	9
45	C	2	75	A	10
48	A	3	80	B	11
50	A	4	87	B	12
57	C	5	90	B	13
58	C	6	95	B	14
60	C	7	98	B	15
62	A	8			

Example II

College A	Rank	College B	Rank	College C	Rank
50	4	80	11	60	7
62	8	95	14	45	2
75	10	98	15	30	1
48	3	87	12	58	6
65	9	90	13	57	5
	R_1=34		R_2=65		R_3=21

Example II

Here $n_1 = 5$, $n_2 = 5$, $n_3 = 5$, $n_2 = 7$, $n = \sum_{i=1}^{k} n_i = 5+5+5 = 15$, $k=3$

$R_1 = 34$, $R_2 = 65$ and $R_3 = 21$

$$\sum_{i=1}^{k} \frac{R_i^2}{n_i} = \frac{34^2}{5} + \frac{65^2}{5} + \frac{21^2}{5} = 1164.40$$

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

$$= \frac{12}{15(15+1)} \times 1164.40 - 3(15+1) = 10.22.$$

The tabled value χ_α^2 for $\alpha=0.05$ and k=3-1 degrees of freedom is 5.99. Since the calculated value $H > \chi_\alpha^2$ test is significant. So we reject H_0 and conclude that the performance students are different for the three colleges.

Example II

Since the test is found significant, we have to determine which colleges are different from each other in median marks for this we conduct Mann Whitney U test. The result presented in the following table. The result indicate that the college A and College B, College B and College C differs significantly. But college A and College C does not differ with each other.

Groups	Compare Colleges	R_i	U_i	U	Result
Group I	College A	15	25.00	0.00	Significant
	College B	40	0.00		
Group II	College A	34	6.00	6.00	Not Significant
	College C	21	19.00		
Group III	College B	40	0.00	0.00	Significant
	College C	15	25.00		
Critical U_α =2					

The difference between Mann Whitney and Kruskal-Wallis Tests

- The major **difference between** the **Mann-Whitney** U and the **Kruskal-Wallis** H is simply that the latter can accommodate more than two groups.
- That is **Mann-Whitney** compare the median or distribution between two groups or two populations.
- **Kruskal-Wallis** compare the median or distribution between three or more groups or populations

Non Parametric Tests

When scaling, group all elements to be scaled. Scale as needed. Use the "Increase Font Size," "Decrease Font Size" buttons or manually change the font

Contact details esjmaths@gmail.com

9847915101, 8089684892

This Completes your syllabus