Application of Statistics in Educational Research I

MPU1034

Nonparametric Statistics*: Kruskal-Wallis Test

Prof. Dr. Mohd Salleh Abu

Dr. Hamidreza Kashefi

main source: Vernoy & Vernoy (1997)

Some Commonly Used Jargons…

- Rank
- Average rank
- Sum of ranks
- Kruskal-Wallis Test (*H* Test)

The Kruskal-Wallis Test

The Needs of Kruskal-Wallis Test

- ☑ ANOVA is applicable for data which take form in either interval or ratio.
- ☑ To conduct a hypothesis testing involving ordinal data gathered from more than two samples, we must employ the Kruskal-Wallis Test.
- ☑ The Kruskal-Wallis Test is also called *H* Test. It is a close counterpart of ANOVA.

Essential Steps in Performing Kruskal-Wallis Test

1. Setting The Hypotheses
2. Ranking the Data
3. Computing the *H* Value
4. Finding Degree of Freedom
5. Finding Critical *H* Value
6. Drawing up a Conclusion

Example

Suppose you conduct an Algebra Readiness Test to 3 different groups of samples, namely Traditional Class, Hands-on Class and Prealgebra Class. Each of these classes comprises of 7 students. The scores obtained by each student are shown in Table 15.18:

Table 15.18 Scores on an Algebra Readiness Test for Samples from Three Classes

Traditional	Hands-on	Prealgebra
95	98	100
94	96	99
89	92	97
84	91	93
83	86	90
81	85	88
80	82	87

Conduct a Kruskal-Wallis test to determine whether there is a significant difference of the Algebra Readiness Test scores among the 3 different groups of students with $\alpha = 0.05$

Setting the Hypothesis

H_o : There is no difference of Algebra Readiness Test scores between the 3 different groups

H_1 : There is at least one group with different Readiness Test scores

Ranking The Data Into Ordinal Scale

- ✓ If the data are already in the ranked form (ordinal type), we can apply the Kruskal-Wallis test straight-away.

If not, we have to rank the data before we can apply the Kruskal-Wallis test.

- ✓ We create ranks by ranking the data. When ranking any set data, we always give the lowest score a rank of 1.

- ✓ In the case which involves scores with multiple occurance – refer to the following example

✓ In the example, the rank are as follows:

Score	Rank
100	21
99	20
98	19
97	18
96	17
95	16
94	15
93	14
92	13
91	12

Score	Rank
90	11
89	10
88	9
87	8
86	7
85	6
84	5
83	4
82	3
81	2
80	1

How Do We Rank Data With Multiple Occurance*?

Suppose we have the following ordinal data:

60 65 67 70 70 70 71 72

We rank the data using the following method:

- Rank 60, 65 and 67 as usual; 1 for 60, 2 for 65 and 3 for 67
- Assign a 'same rank' to all 70s (because they are all tied!). This rank is generated by computing the average of the 'next ranks' of the repeating data, ie. rank 4 (first 70), rank 5 (second 70) and rank 6 (third 7). So, we assign rank

  (4+5+6)/3 = 5

  for each of the 70s
- The next rank should start with 7 (i.e. rank 7 for 71) and rank 8 for 72
- The complete ranking should look like the one shown in the table.

* refer to Vernoy & Vernoy (1997) p 350 for details

Raw Data	Ranked Data
72	8
71	7
70	5
70	5
70	5
67	3
65	2
60	1

(refer to Example)

Table 15.19 Scores on an Algebra Readiness Test and Ranks for Samples from Three Classes

Traditional	R_1	Hands-on	R_2	Prealgebra	R_3
95	16	98	19	100	21
94	15	96	17	99	20
89	10	92	13	97	18
84	5	91	12	93	14
83	4	86	7	90	11
81	2	85	6	88	9
80	1	82	3	87	8
	$\Sigma R_1 = 53$		$\Sigma R_2 = 77$		$\Sigma R_3 = 101$
	$n_1 = 7$		$n_2 = 7$		$n_3 = 7$

Computing The H Value

- ✓ The Kruskal-Wallis test statistic (called *H* value) is computed using the following rules:

$$H = \left[\frac{12}{N_{total} \cdot (N_{total} + 1)} \right] \cdot \left[\frac{(\Sigma R_1)^2}{n_1} + \frac{(\Sigma R_2)^2}{n_2} + \cdots + \frac{(\Sigma R_k)^2}{n_k} \right] - 3 \cdot (N_{total} + 1)$$

The symbols in Formula 15.7 should be familiar to you:

ΣR_1 = the sum of the ranks for sample 1 = 53

ΣR_2 = the sum of the ranks for sample 2 = 77

ΣR_3 = the sum of the ranks for sample 3 = 101

n_1 = the number of scores in sample 1 = 7

n_2 = the number of scores in sample 2 = 7

n_3 = the number of scores in sample 3 = 7

N_{total} = the total number of scores in all the samples = 21

k = the number of groups = 3

- ✓ Substituting these values into the formula, we get

$$H = \left[\frac{12}{21 \cdot (21+1)}\right] \cdot \left[\frac{(53)^2}{7} + \frac{(77)^2}{7} + \frac{(101)^2}{7}\right] - 3 \cdot (21+1)$$

$$= \left(\frac{12}{21 \cdot 22}\right) \cdot \left(\frac{2809}{7} + \frac{5929}{7} + \frac{10{,}201}{7}\right) - 3 \cdot 22$$

$$= \left(\frac{12}{462}\right) \cdot (401.286 + 847 + 1457.286) - 66$$

$$= (0.026 \cdot 2705.572) - 66 = 70.345 - 66$$

$$= 4.345$$

Finding Degree of Freedom (*df*)

✓ The degree of freedom is given by

df = the number of groups - 1

In this case (refer to table),

$df = 3 - 1 = 2$

Finding The Critical Value of H

- ✓ The χ^2 distribution is very close to the distribution of *H*. Therefore, there is no need for a separate table for *H*-values. Instead, we can identify the critical value of *H* (or H_{cv}) using the Table χ .

- ✓ From the Table χ with df = 2, α = 0.05, we would get the critical value of H equal to 5.991

Drawing a Conclusion

Comparing the computed value of H = 4.345 and the critical value of H_{cv} = 5.991, we fail to reject H_o (i.e. there is no significant difference between the samples drawn for the three different classes)

The Chi-Square Distribution
(as an approximation of *H*)