ICCPP-STATISTICS

- Kruskal–Wallis test

Vishal Lohchab

Scientific Assistant of
Prof. Dr. Hans-Werner Gessmann
Director ICCPP International

William Kruskal (1919-2005)

Kruskal–Wallis test

W. Allen Wallis (1912-1998)

Kruskal–Wallis test

Definition

- The Kruskal Wallis test is the non parametric alternative to the One Way ANOVA. Non parametric means that the test doesn't assume your data comes from a particular distribution.

Formula

$$H = \left[\frac{12}{n(n+1)} \sum_{j=1}^{c} \frac{T_j^2}{n_j} \right] - 3(n+1)$$

Where, n = sum of sample sizes for all samples,

c = number of samples,

T_j = sum of ranks in the j^{th} sample,

n_j = size of the j^{th} sample.

Uses

- The H test is used when the assumptions for ANOVA aren't met (like the assumption of normality).

  It is sometimes called the one-way ANOVA on ranks, as the ranks of the data values are used in the test rather than the actual data points.

Uses

- The test determines whether the medians of two or more groups are different. Like most statistical tests, you calculate a test statistic and compare it to a distribution cut-off point.
  The test statistic used in this test is called the H statistic. The hypotheses for the test are:

  - H0: population medians are equal
  - H1: population medians are not equal

Assumptions for the Kruskal Wallis test

Your variables should have:

- One independent variable with two or more levels (independent groups). The test is more commonly used when you have three or more levels.
- Ordinal scale, Ratio Scale or Interval scale dependent variables.

Assumptions for the Kruskal Wallis test

- Your observations should be independent. In other words, there should be no relationship between the members in each group or between groups.
- All groups should have the same shape distributions. Most software (i.e. SPSS, Minitab) will test for this condition as part of the test.

Running the test

- Example

A shoe company wants to know if three groups of workers have different salaries:

Women:	23K, 41K, 54K, 66K, 78K.
Men:	45K, 55K, 60K, 70K, 72K.
Minorities:	18K, 30K, 34K, 40K, 44K.

+
Solution Steps wise

Step 1

Sort the data for all groups/samples into ascending order in one combined set.

- 20K
- 23K
- 30K
- 34K
- 40K
- 41K
- 44K
- 45K
- 54K
- 55K
- 60K
- 66K
- 70K
- 72K
- 90K

Solution Steps wise

Step 2

Assign ranks to the sorted data points. Give tied values the average rank.

- 20K 1
- 23K 2
- 30K 3
- 34K 4
- 40K 5
- 41K 6
- 44K 7
- 45K 8
- 54K 9
- 55K 10
- 60K 11
- 66K 12
- 70K 13
- 72K 14
- 90K 15

Solution Steps wise

Step 3

Add up the different ranks for each group/sample.

Women: 23K, 41K, 54K, 66K, 90K = 2 + 6 + 9 + 12 + 15 = 44

Men: 45K, 55K, 60K, 70K, 72K = 8 + 10 + 11 + 13 + 14 = 56.

Minorities: 20K, 30K, 34K, 40K, 44K = 1 + 3 + 4 + 5 + 7 = 20.

Solution Steps wise

Step 4.1

Calculate the H statistic

$$H = \left[\frac{12}{n(n+1)} \sum_{j=1}^{c} \frac{T_j^2}{n_j} \right] - 3(n+1)$$

Where, n = sum of sample sizes for all samples,

c = number of samples,

T_j = sum of ranks in the j^{th} sample,

n_j = size of the j^{th} sample.

Solution Steps wise

Step 4.2

Calculate the H statistic:

$$H = \left[\frac{12}{15(15+1)} \left[\frac{44^2}{5} + \frac{56^2}{5} + \frac{20^2}{5} \right] - 3(15+1) \right]$$

$$H = 6.72$$

+

Solution Steps wise

Step 5

Find the critical chi-square value, with c-1 degrees of freedom. For 3 – 1 degrees of freedom and an alpha level of .05, the critical chi square value is 5.9915.

Solution Steps wise

Step 6.1

Compare the H value from Step 4 to the critical chi-square value from Step 5.

If the critical chi-square value is less than the H statistic, reject the null hypothesis that the medians are equal.

Solution Steps wise

Step 6.2

If the chi-square value is not less than the H statistic, there is not enough evidence to suggest that the medians are unequal.

In this case, 5.9915 is less than 6.72, so you can reject the null hypothesis.

References

Beyer, W. H. CRC Standard Mathematical Tables, 31st ed. Boca Raton, FL: CRC Press, pp. 536 and 571, 2002.

Agresti A. (1990) Categorical Data Analysis. John Wiley and Sons, New York.

Gonick, L. (1993). The Cartoon Guide to Statistics. HarperPerennial.

Wheelan, C. (2014). Naked Statistics. W. W. Norton & Company.

Stephanie Glen. "Welcome to Statistics How To!“
From StatisticsHowTo.com: Elementary Statistics for the rest of us!
https://www.statisticshowto.com/