Applied Mechanics and Materials Vol. 611 (2014) pp 115-120
© (2014) Trans Tech Publications, Switzerland
doi:10.4028/www.scientific.net/AMM.611.115

Submitted: 2014-05-27
Accepted: 2014-06-05
Online: 2014-08-11

Methodology and Application of the Kruskal-Wallis Test

Eva Ostertagová[1,a], Oskar Ostertag[2,b] and Jozef Kováč[3,c*]

[1]Faculty of Electrical Engineering and Informatics, TU Košice, Letná 9, Košice, Slovakia

[2, 3]Faculty of Mechanical Engineering, Technical University of Košice, Letná 9, Košice, Slovakia

[a]eva.ostertagova@tuke.sk, [b]oskar.ostertag@tuke.sk, [c]jozef.kovac@tuke.sk

Keywords: Kruskal-Wallis test, nonparametric multiple comparisons, Conover-Inman test, MATLAB

Abstract. This paper describes the methodology and application of the very popular nonparametric test which is a rank based test named as Kruskal-Wallis. This test is useful as a general nonparametric test for comparing more than two independent samples. It can be used to test whether such samples come from the same distribution. This test is powerful alternative to the one-way analysis of variance. Nonparametric ANOVA has no assumption of normality of random error but the independence of random error is required. If the Kruskal-Wallis statistic is significant, the nonparametric multiple comparison tests are useful methods for further analysis. The statistical analysis of the application data in this paper was performed with software MATLAB.

Introduction

Nonparametric methods require less stringent assumptions than do their parametric counterparts; on the other hand, they also use less information from the data. This makes the nonparametric tests somewhat less powerful than the corresponding parametric tests for the same situations, when the assumptions of the parametric tests are met. When the assumptions of the parametric tests are not met, the nonparametric tests are the ones we should use [1].

The ANOVA F-test is used to test the equality more than two population means. However, the F-test has some assumptions that are frequently ignored and often violated when used in real world applications. These assumptions include that the data in each group come from a normal distribution, that the population variances in each group are equal (homoscedasticity) and that the data are independent of one another. If these assumptions are met, than the F-test of ANOVA is a powerful method for determining whether several population means are equal. When the means of k populations are compared and it is known that the populations do not have equal variances or that the populations are not normal, the Kruskal-Wallis nonparametric test is used [2,3].

The Kruskal-Wallis test (Kruskal and Wallis 1952, 1953) is the nonparametric equivalent of a one-way ANOVA and is used for testing whether samples originate from the same distribution. This test is basically an extension of the Wilcoxon-Mann-Whitney two sample test (Wilcoxon 1945, Mann and Whitney 1947) for more than two independent samples.

The Kruskal-Wallis test does not make assumptions about normality. However, it assumes that the observations in each group come from populations with the same shape of distribution and that the samples are random and independent. The test statistic for one-way analysis of variance is calculated as the ratio of the treatment sum of squares to the residual sum of squares. The Kruskal-Wallis test uses the same method but, as with many nonparametric tests, the ranks of the data are used in place of the raw data [1].

The popularity of Kruskal-Wallis test may be attributed to its usefulness in a variety of disciplines such as engineering and manufacturing applications, medicine, biology, psychology and education.

The Kruskal-Wallis test – a nonparametric alternative to one-way ANOVA

When performing the Kruskal-Wallis test the following assumptions are required:

- The continuous distributions for the test variable are exactly the same (except their medians) for the different populations.
- The cases represent random samples from the populations, and the scores on the test variable are independent of each other.

Assume that the data $x_{11}, x_{12}, x_{13}, \dots, x_{1n_1}$ are sample from population 1, $x_{21}, x_{22}, x_{23}, \dots, x_{2n_2}$ are sample from population 2, … , $x_{k1}, x_{k2}, x_{k3}, \dots, x_{kn_k}$ are sample from population *k*. Let x_{ij}, $i = 1, 2, \dots, k$ and $j = 1, 2, \dots, n_i$ denote the data from i^{th} group (level) and j^{th} observation, and $F_i(x)$ denote the continuous distributions of the x_{ij}.

Further, suppose that this independent random samples of sizes $n_1, n_2, \dots, n_k$, are drawn from *k* continuous (not necessarily normal) populations.

We will be interested in testing the null hypothesis

$$H_0: F_1(x) = F_2(x) = \cdots = F_k(x), \text{ for all } x \text{ (distributions for all } k \text{ populations are all the same)}, \quad (1)$$

against the alternative hypothesis

$$H_1: \exists\, 1 \le i, l \le k: F_i(x) \ne F_l(x) \text{ (at least two population distributions differ in location).} \quad (2)$$

These hypotheses can be made more specific (e.g. null hypothesis H_0: population medians are the same).

We rank all $N = \sum_{i=1}^{k} n_i$ observations from smallest to largest, without regard to which sample they come from, and assign the smallest observation rank 1, the next smallest rank 2, … , and the largest observation rank *N*.

Let R_{ij} be the rank of the data point x_{ij}, and define R_i to be the sum of the ranks in the i^{th} sample. That is $R_i = \sum_{j=1}^{n_i} R_{ij}$, and then denote each sample mean by $\bar{R}_i = \frac{R_i}{n_i}$.

Let $\bar{R}$ represent the overall mean, and because is $\sum_{i=1}^{k} R_i = \frac{N(N+1)}{2}$, then $\bar{R} = \frac{\sum_{i=1}^{k} R_i}{N} = \frac{N+1}{2}$.

The Kruskal-Wallis test statistic measures the degree to which the actual observed mean ranks $\bar{R}_i$ differ from their expected value $(N+1)/2$. If this difference is large, the null hypothesis H_0 is rejected.

The test statistic to be computed if there are no ties (that is, if no two observations are equal) is [4,5]:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} n_i \left(\bar{R}_i - \bar{R}\right)^2 = \frac{12}{N(N+1)} \sum_{i=1}^{k} n_i \left(\frac{R_i}{n_i} - \frac{N+1}{2}\right)^2. \quad (3)$$

The coefficient $12/N(N+1)$ is a suitable normalization factor [6].

The statistic *H* may be rewritten in other equivalent form as

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1). \quad (4)$$

Whenever H_0 is true and either
- $k = 3$ and $n_i \geq 6$ for $i = 1, 2, 3$
- $k > 3$ and $n_i \geq 5$ for $i = 1, 2, \dots, k$

the distribution of the test statistic H is well approximated by the chi-square distribution with $k-1$ degrees of freedom [2]. For very small samples ($n_i < 5$) there are special tables for the exact distribution of H statistic under the null hypothesis.

We reject the null hypothesis on the right-hand tail of the chi-square distribution. That is, the null hypothesis H_0 is rejected on significance level α, when

$$H > \chi^2_{1-\alpha,\, k-1}\,, \tag{5}$$

where $\chi^2_{1-\alpha,\, k-1}$ is $(1-\alpha)$-quantile of the chi-square distribution with $k-1$ degrees of freedom [1,4].

If there are ties, each observation is given the mean of the ranks for which it is tied. H statistic as computed from (4) is then divided by the correction factor

$$f^* = 1 - \frac{\sum_{i=1}^{m}\left(t_i^3 - t_i\right)}{N^3 - N}\,, \tag{6}$$

where t_i is the number of ties in i^{h} group of m tied groups [4].

Then the test statistic corrected for ties is

$$H^* = \frac{H}{f^*}\,. \tag{7}$$

The test statistic H^* has also an asymptotic chi-square distribution based on $k-1$ degrees of freedom. The effect of correcting for ties is to increase the value of H thus to make the result more significant than it would have been had no correction been made.

Nonparametric multiple comparisons

When the Kruskal-Wallis test leads to significant results, then at least one of the samples is different from the other samples. The test does not identify where the differences occur or how many differences actually occur. Therefore, a test procedure for making pair-wise comparisons is needed. This procedure tests the null hypothesis that two samples I and J ($I, J = 1, 2, \dots, k$, for $I \neq J$) are from identical population against the alternative hypothesis that two samples come from different populations.

The null and alternative hypotheses are of the form

$$H_0:\ F_I(x) = F_J(x)\,,\ H_1:\ F_I(x) \neq F_J(x)\,. \tag{8}$$

Most multiple comparison procedures for population means can be applied in the nonparametric case, using mean ranks instead of sample means, and a nonparametric statistic.

A common procedure used with Kruskal-Wallis method is the Conover-Inman post hoc procedure. The Conover-Inman (1999) procedure is simply Fisher LSD (least significant difference) method performed on ranks. Samples I and J are significantly different at the significance level α if the absolute value of the difference between their mean ranks is greater than the least significant difference, i.e. if the following inequality is satisfied

$$\left|\overline{R}_I - \overline{R}_J\right| > t_{1-\alpha/2,\, N-k} \cdot \sqrt{\frac{N(N+1)}{12} \cdot \frac{N-1-H}{N-k} \cdot \left(\frac{1}{n_I} + \frac{1}{n_J}\right)}\,, \tag{9}$$

where $t_{1-\alpha/2,\,N-k}$ is $(1-\alpha/2)$-quantile of the Student t-probability distribution with $N-k$ degrees of freedom. If there are ties, this can be calculating in corrected form [7,8,9]:

$$\left|\overline{R}_I-\overline{R}_J\right|>t_{1-\alpha/2,\,N-k}\cdot\sqrt{\left(\frac{N(N+1)}{12}-\frac{\sum_{i=1}^{m}\left(t_i^3-t_i\right)}{12(N-1)}\right)\cdot\frac{N-1-H^*}{N-k}\cdot\left(\frac{1}{n_I}+\frac{1}{n_J}\right)}\,. \tag{10}$$

Example from technical practice

What follows is an example of Kruskal-Wallis procedure. We are interested in the effect on tube conductivity of five different type of coating for cathode ray tubes in a telecommunications system display device. The data from this experiment are in the Table 1. Using a 5 % significance level, is there any difference due to coating type?

Table 1 Experiment data

Type 1	143	141	150	146	145
Type 2	150	149	137	134	152
Type 3	134	133	132	127	128
Type 4	129	127	132	129	130
Type 5	147	148	144	142	143

One of the first steps in using the parametric one-way ANOVA is to test the assumption of the normality. The normality of data distribution was verified with Shapiro-Wilk goodness of fit test. The null hypothesis is that there is no significant departure from normality for each of the groups. The alternative hypothesis is that there is a significant departure from normality. Shapiro-Wilk test was implemented using software MATLAB with function *swtest* and a significance level of 0.05 was used. We have received these results: $p=0.9212$ for coating type 1, $p=0.1509$ for type 2, $p=0.3317$ for type 3, $p=0.8258$ for type 4, and $p=0.5012$ for type 5. Since *p*-value was larger than given significance level 0.05 for each of our five groups, we would conclude that all of the treatments are normally distributed.

After it, we tested equality of variances. The null hypothesis is $H_0:\ \sigma_1^2=\sigma_2^2=\cdots=\sigma_5^2$, and the alternative hypothesis is $H_1:\ \exists\, 1\le i,l\le 5:\ \sigma_i^2\ne\sigma_l^2$. We have run Bartlett test in MATLAB with using function *vartestn* and significance level 0.05. We have received the *p*-value 0.0312. Since *p*-value was less than given significance level 0.05 for this problem, we rejected the null hypothesis, which further demonstrated that parametric one-way ANOVA was inappropriate.

We therefore used a nonparametric ANOVA, the Kruskal-Wallis test, to determine whether differences between treatments existed. We want to test the null hypothesis that population medians are equal, versus the alternative that there is a difference between at least two of them.

There are $k=5$ populations, the sample sizes are equal $n_i=5$. Each of the $N=25$ observations was replaced by a rank relative to all the observations in all of the samples. Where several values were tied (i.e. identical), each was given the mean of their ranks. The ranks are shown in parenthesis in the Table 2.

Table 2 Experiment data and their ranks

Type 1	143 (15.5)	141 (13)	150 (23.5)	146 (19)	145 (18)
Type 2	150 (23.5)	149 (22)	137 (12)	134 (10.5)	152 (25)
Type 3	134 (10.5)	133 (9)	132 (7.5)	127 (1.5)	128 (3)
Type 4	129 (4.5)	127 (1.5)	132 (7.5)	129 (4.5)	130 (6)
Type 5	147 (20)	148 (21)	144 (17)	142 (14)	143 (15.5)

For the data in Table 2, the sums of ranks for each ward are $R_1 = 89$, $R_2 = 93$, $R_3 = 31.5$, $R_4 = 24$, $R_5 = 87.5$, and means of ranks are $\overline{R}_1 = 17.8$, $\overline{R}_2 = 18.6$, $\overline{R}_3 = 6.3$, $\overline{R}_4 = 4.8$, $\overline{R}_5 = 17.5$. There are six pairs of tied observations.

Using the formulas (4), (6), and (7) were obtained the Kruskal-Wallis *H*-statistic, the correction factor for ties, corrected value of the *H*-statistic for ties, with these results $H = 17.2412$, $f^* = 0.9977$, $H^* = 17.2811$. Since the obtained H^* value was larger than critical point of the chi-square distribution $\chi^2_{0.95,\,4} = 9.4877$, the null hypothesis H_0 was rejected on significance level $\alpha = 0.05$. We found a significant difference between treatments using the Kruskal-Wallis test.

For Kruskal-Wallis test procedure in MATLAB we can use function *kruskalwallis*. The output for our example gives the *p*-value 0.0017, indicating that at least one of the coating types is different from others. Boxplots for the five different groups are shown in Fig. 1.

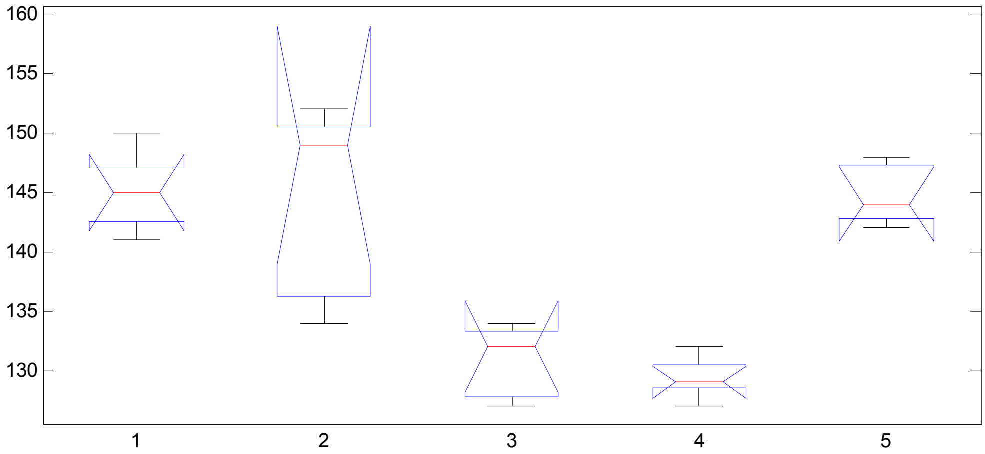

Fig. 1 Boxplots for the five types of coating

When the null hypothesis that population medians are equal is rejected, we want to know where the difference among the medians is. To determine which pairs of medians are significantly different, and which are not, we are used the Conover-Inman test procedure (see formula 10). The statistical outputs are shown in the Table 3.

Table 3 Results using Conover-Inman test

Pairs *I*, *J*	Difference $\lvert \overline{R}_I - \overline{R}_J \rvert$	Least significant difference
1, 2	0.8	5.6213
1, 3 ∗	11.5	5.6213
1, 4 ∗	13	5.6213
1, 5	0.3	5.6213
2, 3 ∗	12.3	5.6213
2, 4 ∗	13.8	5.6213
2, 5	1.1	5.6213
3, 4	1.5	5.6213
3, 5 ∗	11.2	5.6213
4, 5 ∗	12.7	5.6213

It may be concluded that the medians between treatment groups with sign ∗ in the Table 3 are different.

Conclusion

The two most commonly used tests that are available to the statistical analyst are one-way ANOVA and the Kruskal-Wallis test.

Unlike the parametric ANOVA, the Kruskal-Wallis nonparametric one-way ANOVA does not require the fulfilment of assumptions of normal distribution, interval data and homogeneity of group variance. This test is a more flexible, convenient, easy to use and powerful technique similar to a parametric one-way ANOVA [10].

Acknowledgement

This article was created by implementation grant project VEGA no. 1/1205/12.

References

[1] A.D. Aczel, Complete Business Statistics, Irwin, 1989.

[2] D.C. Montgomery, G.C. Runger, Applied Statistics and Probability for Engineers, John Wiley & Sons, New York, 2003.

[3] E. Ostertagová, O. Ostertag, Methodology and Application of One-way ANOVA, in: American Journal of Mechanical Engineering, 2013, Vol. 1, No. 7, pp. 256-261.

[4] W.H. Kruskal, W.A. Wallis, Use of Ranks in One-Criterion Variance Analysis, in: Journal of the American Statistical Association, 1952, Vol. 47, No. 260, pp. 583-621.

[5] J.D. Gibbons, S. Chakraborti, Nonparametric Statistical Inference, fourth ed., Taylor & Francis e-Library, 2005.

[6] E.B. Manoukian, Mathematical nonparametric statistics, Gordon and Breach Science Publishers, 1986.

[7] V. Bewick, L. Cheek, J. Ball, Statistics review 10: Further nonparametric methods, Critical Care, 2004, Vol. 8, No. 3, pp. 196-199.

[8] W.J. Conover, Practical nonparametric statistic, third ed., John Wiley & Sons, New York, 1999.

[9] http://www.statsdirect.com/help/default.htm#nonparametric_methods/kruskal_wallis.htm

[10] D. Israel, Data Analysis in Business Research, first ed., Sage Publications Ltd, 2008.

Applied Mechanics and Mechatronics
10.4028/www.scientific.net/AMM.611

Methodology and Application of the Kruskal-Wallis Test
10.4028/www.scientific.net/AMM.611.115