Dariusz Parys*

KRUSKAL-WALLIS TEST IN MULTIPLE COMPARISONS

Abstract

In this paper we show that the Kruskal-Wallis test can be transform to quadratic form among the Mann-Whitney or Kendall $\tau$ au concordance measures between pairs of treatments. A multiple comparisons procedure based on patterns of transitive ordering among treatments is implement. We also consider the circularity and non-transitive effects.

Key words: multiple comparisons, circularity, transitive and non-transitive effects.

1. Introduction

Consider $k$ samples of independent observations, where the $i$th sample of $n_i$ observations $\{x_{i1}, x_{i2}, ..., x_{in_i}\}$ is drawn from a population with cumulative distribution function (cdf) $F_i$, representing the effect of the $i$th “treatment”. Let $R_i$ be the average rank of the $i$th sample in the overall pooled sample of $N$ observations. The Kruskal-Wallis test statistic of $H_0$: all $\{F_i\}$ equal is

$$KW = \frac{12}{N(N+1)} \sum_{i=1}^{k} n_i \left( R_i - \frac{1}{2}(N+1) \right)^2$$

(1)

whose null distribution is approximately $\chi^2_{k-1}$.

Now consider Mann-Whitney statistics, which are the concordance measures used in the definition of Kendall’s tau, between each pair of samples. Let

$$T_{ij} = \sum_{b=1}^{n_j} \sum_{a=1}^{n_i} \text{sgn}(x_{jb} - x_{ia})$$

* Ph.D., Department of Statistical Methods, University of Łódź.
the difference between the number of times a \( j \)-th sample observation exceeds, or is less than, an \( i \)-th sample observation. A large value of \( T_{ij} \) signifies that treatment \( j \) observations tend to exceed those from treatment \( i \).

When a location shift model is appropriate,

\[
p_{ij} = \Pr(x_j > x_i) = \int F(t + \Delta_{ij})dF(t)
\]

where \( F \) denotes the model distribution and \( \Delta_{ij} \) denotes the shift between the distributions. Now if \( \Delta_{ij} \geq 0 \) for \( i < j \) then \( p_{13} \geq \max\{p_{12}, p_{23}\} \). This means that if \( T_{12}, T_{23} \) and \( T_{23} \) are large, then \( T_{13} \) is also large. Hence, we have transitivity of effects.

We will consider the case \( k = 3 \). It is shown that \( KW \) is a quadratic form among \( T_{12}, T_{23}, T_{13} \), but there is a single degree of freedom left over, attributable to a circularity contrast, uncorrelated with \( KW \). It may be that \( T_{12} \) and \( T_{23} \) are large but \( T_{13} \) is small suggesting the non-transitive effects described as \( A_1 > A_2 > A_3 > A_1 \), where \( A_i \) stands for treatment \( i \). Therefore within the full set of concordance measures there is information about circularity as well as about \( KW \), the latter being regarded as assessing transitive effects, corresponding to a linear ordering among treatments.

2. The case of three treatments

From the concordance measures \( \{T_{ij}\} \) further contrasts can be defined to detect certain ordering between treatments. Let \( T_1 = T_{12} + T_{31} \), \( T_2 = T_{12} + T_{32} \) and \( T_3 = T_{13} + T_{23} \). Thus noting that \( T_{ij} = -T_{ji} \), large values of \( T_1 \) indicate that \( A_1 > A_2, A_3 \), while large values of \( T_1 - T_3 = T_{32} + 2T_{31} + T_{32} \) are indications of \( A_1 > A_2 > A_3 \). Note that \( T_1 + T_2 + T_3 = 0 \).

In the following Theorem 1, we show that the Kruskal-Wallis statistic (1) can be written in terms of \( T_1, T_2 \) and \( T_3 \), and hence in terms of \( T = (T_{11}, T_{23}, T_{31}) \). This means that ranks of the combined data can be replaced by pairwise rankings.

**Theorem 1.** Another expression for the Kruskal–Wallis test statistics

\[
KW = \frac{3}{N(N+1)} \left( \frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \frac{T_3^2}{n_3} \right)
\]

Each contrast \( T_i \) measures the tendency of treatment \( i \) to have higher responses than other treatments, and \( KW \) captures this effects over all treatments.
Thus \( KW \) is magnified by a definite linear or transitive ordering among the treatments.

We next provide the covariance matrix \( T = (T_{12}, T_{23}, T_{31}) \) and use it to define a natural quadratic form in \( T \). This quadratic form includes the information in the pairwise Mann-Whitney statistics for testing \( H_0 \). We further show that the Kruskal-Wallis statistic is only part of this quadratic form and the remainder is a quadratic form that is sensitive to intransitivities in the data.

Let \( T = (T_{12}, T_{23}, T_{31}) \). It is easy to verify that the covariance matrix of \( T \) is

\[
V_T = \begin{pmatrix}
n_1n_2(n_1+n_2+1) & -n_1n_2n_3 & -n_2n_3 \\
-n_1n_2 & n_2(n_1+n_1+1) & -n_1n_3 \\
-n_1n_2 & -n_2n_3 & n_3(n_1+n_1+1)
\end{pmatrix}.
\]

**Theorem 2.** The expression for \( T^TV_T^{-1}T \) is

\[
Q_3 = \frac{3}{N+1} \left( \frac{T_{12}^2(1+n_3)}{n_1n_2} + \frac{T_{23}^2(1+n_1)}{n_2n_3} + \frac{T_{31}^2(1+n_2)}{n_3n_1} + \frac{2T_{12}T_{23}}{n_1n_2} + \frac{2T_{12}T_{31}}{n_1n_3} + \frac{2T_{23}T_{31}}{n_2n_3} \right).
\]

A different type of effect is measured by the circularity contrast

\[
C_{123} = \frac{T_{12}}{n_1n_2} \frac{T_{23}}{n_2n_3} \frac{T_{31}}{n_3n_1},
\]

because large values of \( C_{123} \) indicate a tendency for \( A_1 < A_2 < A_3 < A_4 \), a circular or non-transitive effect.

Under \( H_0 \), \( C_{123} \) is uncorrelated with \( T_1, T_2 \) and \( T_3 \); \( \text{var}(C_{123}) = N/(3n_1n_2n_3) \).

**Theorem 3.** A transitive/non-transitive decomposition of \( Q_3 \) is given by

\[
Q_3 = KW + Q_C, \quad \text{where} \quad Q_C = \frac{3n_1n_2n_3}{N} C_{123}^2.
\]

**Corollary to Theorem 3.** \( KW \) and \( Q_C \) are asymptotically independent as \( N \to \infty \), and under \( H_0 \)

\[
KW \overset{d}{\to} \chi^2_2 \quad \text{as} \quad N \to \infty, \quad \text{provided} \quad \lim_{N \to \infty} \frac{n_i}{N} = \lambda_i > 0, \quad \text{for} \quad i = 1, 2, 3
\]
3. Efron dice

This section discusses the questions:
- What do non-transitive samples look like?
- How might we simulate samples from populations for which $KW$ is not significant but for which there are significant circularities?

One approach to generating non-transitive samples can be developed from considering a set of dice first proposed by Efron and described by Gardner (1970). We refer to these dice as Efron dice. Figures 1–3 show some examples of non-transitive Efron dice. Termy & Foster (1976) provide an algorithm for their construction.

Consider a simulation of non-transitive samples based on Efron dice in Figure 1. Let $f_a$ denote the $N(a, 1)$ density function. For a die with faces $i_1, \ldots, i_6$ take the corresponding distribution to have density $\frac{1}{6} \sum_{j=1}^{6} f_{i_j}$, i.e. a mixture of unit variance normal distributions whose means are the die-face markings.

In a simulation experiment, 20 observations were generated from each of the three distributions for the dice in Figure 1, and the values of the Kruskal-Wallis and the circularity quadratic forms $KW$ and $QC$ were calculated. The experiment was repeated 100 times, and the average values of $KW$ and $QC$ were 2.16 and 4.28 respectively. The approximate null distribution of $KW$ is $\chi^2_2$, with mean 2, and the value $KW = 2.16$ is not significant. But the null mean of $QC$ is 1, so the value $QC = 4.28$ appears to be inflated by the presence of circularity effects. When the whole experiment was repeated with 30 observations from each die-distribution, the average $KW$ and $QC$ values were 1.93 and 6.11 respectively, again exhibiting non-significance of the transitive $KW$, and the apparent strong presence of circularity effects in $QC$.

Thus the generated samples show no overall indication of any statistical significance for relative shifts in the populations, with non-significant Kruskal-Wallis tests. On the other hand, the statistics measuring non-transitivity are inflated above mean values and suggest circularity effects.

If the samples were expressions of treatment effects, then the effects would arise from mixture distributions rather than location shifts from a control population. This sort of effect arises when patients react differently to a drug. Some may get a positive effect, some a negative effect (reaction), and some may be unaffected; see Boos and Brownie (1991) for an example.
Fig. 1. There are non-transitive dice A, B, and C. Let $A > B$ denote that die A beats die B in a single toss of the dice. Then $\Pr(A > B) = \Pr(B > C) = \Pr(C > A) = \frac{5}{9}$.

Fig. 2. Four non-transitive dice A, B, C, and D. $\Pr(A > B) = \Pr(B > C) = \Pr(C > D) = \Pr(D > A) = \frac{2}{3}$, $\Pr(C > A) = \frac{5}{9}$, $\Pr(D > B) = \frac{1}{2}$. These dice have both a 4-cycle and a 3-cycle.

Fig. 3. Four non-transitive dice A, B, C, and D. $\Pr(A > B) = \frac{21}{36}$, $\Pr(B > C) = \frac{25}{36}$, $\Pr(C > D) = \frac{21}{36}$, $\Pr(D > A) = \frac{18}{36}$, $\Pr(C > A) = \Pr(D > B) = \frac{21}{36}$. These are two 3-cycle, an eddy, but not a 5-cycle.

4. The case of general $k$

As described in Theorem 3, the full-rank quadratic form with 3 degrees of freedom (df) is decomposable into the sum of two asymptotically independent quadratic forms, one of Kruskal-Wallis type with 2 df, the other a single (circulant) with 1 df, and these two together exhaust all available degrees of freedom.
Within the Kruskal-Willis test, it is often possible to make statements supporting specific orderings among the populations. The full $KW$ quadratic form is given by (1) or (2), is asymptotically $\chi^2$, and, if significantly large, indicates a most significant linear combination among the $KW$ contrasts $T_1, T_2, T_3$ in the Scheffé sense. However, there are particular simpler contrasts which have definite meaning in terms of supporting ordering statements among $A_1, A_2, A_3$ as alternatives to $H_0$ (see Table 1).

<table>
<thead>
<tr>
<th>Large values of $T_1$</th>
<th>Large values of $T_1 - T_2$</th>
<th>Large values of $T_1 - T_3$</th>
<th>Large values of $T_2 - T_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1 &gt; A_2, A_3$</td>
<td>$A_1 &gt; A_3 &gt; A_2$</td>
<td>$A_1 &gt; A_2 &gt; A_3$</td>
<td>$A_1 &gt; A_3 &gt; A_2$</td>
</tr>
<tr>
<td>$-T_1$</td>
<td>$A_1 &lt; A_2, A_3$</td>
<td>$A_1 &gt; A_3 &gt; A_2$</td>
<td>$A_1 &gt; A_3 &gt; A_2$</td>
</tr>
<tr>
<td>$T_2$</td>
<td>$A_2 &gt; A_1, A_3$</td>
<td>$A_1 &gt; A_3 &gt; A_2$</td>
<td>$A_1 &gt; A_3 &gt; A_2$</td>
</tr>
<tr>
<td>$-T_2$</td>
<td>$A_2 &lt; A_1, A_3$</td>
<td>$A_1 &gt; A_3 &gt; A_2$</td>
<td>$A_1 &gt; A_3 &gt; A_2$</td>
</tr>
<tr>
<td>$T_3$</td>
<td>$A_3 &gt; A_2, A_1$</td>
<td>$A_2 &gt; A_1 &gt; A_3$</td>
<td>$A_2 &gt; A_1 &gt; A_3$</td>
</tr>
<tr>
<td>$-T_3$</td>
<td>$A_3 &lt; A_2, A_1$</td>
<td>$A_2 &gt; A_1 &gt; A_3$</td>
<td>$A_2 &gt; A_1 &gt; A_3$</td>
</tr>
</tbody>
</table>

Source: own study.

If $H_0$ is to be rejected, it is useful to make post-hoc statements about the ordering of responses from the $k$ different treatments $A_1, ..., A_k$. Some of these orderings are transitive, like $A_1 > A_2 > ... > A_k$, detected by the Kruskal-Wallis test. Others are non-transitive circulants like $A_1 > A_2 > ... > A_k > A_1$, not detected by a Kruskal-Wallis test. There are $\binom{k}{2} = \frac{1}{2}k(k-1)$ different concordance measures $\{T_{ij}\}$, but the Kruskal-Wallis test uses only $k - 1$ df. This suggests the residue $\frac{1}{2}(k-1)(k-2)$ df are available for circularities. For $k = 3$, this residue = 1, corresponding to the single circulant $C_{123}$.

Call a circulant $C_{h,i,j}$ a 3-cycle or primary circulant. An $r$-cycle has the form

$$C_{k_1k_2...k_r} = \frac{T_{k_1}}{n_i n_{i_2}} + \frac{T_{k_2}}{n_i n_{i_3}} + \ldots + \frac{T_{k_r}}{n_i n_{i_1}}$$

All $r$-cycles can be expressed in terms of primary circulants, for example $C_{1234} = C_{123} + C_{341}$. In comparison with circulants, contrasts like $T_1, T_2$ are the building blocks for transitive orderings. Call $\sum_{\{\text{some } j\neq i\}} T_{ji}$ a direct sum for
treatment \( i \), where large values indicate that treatment \( i \) has response levels exceeding the other treatments in the sum.

Consider the idea of Kruskal-Wallis contrasts, which are direct sums including all other treatments. The Kruskal-Wallis contrast for treatment \( i \) is

\[
T_i = \sum_{(\text{all} \neq i)} T_{ji}.
\]

Large values for \( T_i \) signify that responses for treatment \( i \) generally exceed those of all other treatments.

A useful graph theory representation identifies each treatment \( A_1, A_2, \ldots \) with a vertex, and connects every pair \( A_i, A_j \) with an edge. Call each triangle with vertices \( A_i, A_j, A_k \) a surface (Figure 4). Each \( T_{h,i} \) can be envisaged as measuring a “flow rate” along the edge \( A_i A_j \). If \( T_{h,i} \) is positive, it signifies that

\[
p_{i,j} > \frac{1}{2},
\]

where \( p_{i,j} \) is the probability that a \( j \)th treatment observation exceeds an \( i \)th treatment observation.
A direct sum is the sum of some or all $T_j$ terms along edges towards a single vertex $A_i$. A primary circulant is the sum of three $T_j$ terms around the edges of a surface; see Figure 5. When $k = 4$, a 4-cycle is a tour around the edges connecting four vertices; see Figure 6. An eddy consists of two adjacent 3-cycles, having opposite rotational direction, with a reinforced “flow” along one common edge; see Figure 7.

**Theorem 5.** Define $K$, the Kruskal-Wallis vector of normalized Kruskal-Wallis contrasts by $K = \left( T_1 / \sqrt{n_1}, ..., T_k / \sqrt{n_k} \right)$. Then its covariance matrix is

$$V_k = \frac{1}{3} N(N+1)(I - N^{-1} uu^T)$$

where $u = (\sqrt{n_1}, ..., \sqrt{n_k})$.

**Corollary to Theorem 5.** As $N$ and all $\{n_i\} \to \infty$, if condition (4) holds, then under $H_0$,

$$\frac{3}{N(N+1)} K^T K = \frac{3}{N(N+1)} \sum_{i=1}^{k} \frac{T_i^2}{n_i} \rightarrow \chi^2_{k-1}$$

**5. An example**

The following example is taken from Anderson (2000). The data are from a survey of expatriate employees, returning from a period abroad, asking their opinions of the adequacy of the preparation, training and support that they received. The responses came from three groups of employees: $A_1$, private enterprise; $A_2$, government; and $A_3$, religious organizations. The details of a multiple comparisons ANOVA, using the formulas and procedures outlined in the paper are as follows and also in Table 4.

| Table 4 |
| Kruskal-Wallis multiple comparisons analysis (the $P$-values are obtained by comparing standardized values with $\chi^2_2$ values) |

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Value</th>
<th>Null sd</th>
<th>Standardized value</th>
<th>$P$-value</th>
<th>Supports</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-T_1$</td>
<td>1226</td>
<td>384</td>
<td>10.18</td>
<td>0.006</td>
<td>$A_1 &lt; A_2, A_3$ ($A_1$ worst)</td>
</tr>
<tr>
<td>$T_3 - T_1$</td>
<td>2014</td>
<td>641</td>
<td>9.87</td>
<td>0.007</td>
<td>$A_1 &lt; A_2 &lt; A_3$</td>
</tr>
<tr>
<td>$T_3$</td>
<td>788</td>
<td>357</td>
<td>4.88</td>
<td>0.087</td>
<td>$A_1 &lt; A_3 &lt; A_2$</td>
</tr>
<tr>
<td>$T_1 - T_2$</td>
<td>1664</td>
<td>668</td>
<td>6.21</td>
<td>0.045</td>
<td>$A_1 &lt; A_2 &lt; A_3$ ($A_3$ best)</td>
</tr>
</tbody>
</table>

**Source:** own study.
\[ n_1 = 47, \; n_2 = 41, \; n_3 = 35 \]

\[ T_{12} = 623, \; T_{23} = 185, \; T_{31} = -603 \]

\[ KW = 10.70, \; p\text{-value} = 0.005, \text{using} \chi^2_2, \; Q_c = 12.07 \]

The interpretation of the transitive, \( KW \) part of the analysis is straightforward. There is strong evidence suggesting that \( A_1 < A_2 < A_3 \), and in particular that \( A_1 \) is the “worst” treatment. Both of the orderings \( A_1 < A_3 < A_2 \) and \( A_1 < A_2 < A_3 \) represent departures from the pull hypothesis in different, though similar, directions, but the evidence suggesting \( A_1 < A_2 < A_3 \) is considerably stronger than for \( A_1 < A_3 < A_2 \).

Another interesting aspect of the analysis is the circularity contrast \( C_{123} = 0.0857 \), whose normalized value is 3.47. If this had been a normalized \( z \)-value, it would be highly significant with a two-sided \( P < 0.001 \), and even with the non-normal limit distribution it is still significant with estimated \( P < 0.001 \). The question of how to interpret circularities is a separate issue which deserves a more thorough discussion, along with the asymptotic theory. Note however that transitive \( KW \) effects can interfere with circularity effects. For example, a rank ordering \( AAAAAABBBBBCCCCC \) would be interpreted by most statisticians as arising from pronounced transitivity \( A < B < C \), yet the circulant statistic \( C_{123} = 1 \) is significant; \( P < 0.001 \) from permutation testing.

Hence, as mentioned earlier, we recommend that the test for circularity be performed on the residuals only after removing the transitive location effects. When this is done for the example in this section, we obtain \( C_{123} = 5.1 \), with permutation \( P \)-value < 0.001. This constitutes strong evidence against a simple location shift model.

**References**


Dariusz Parys

**Tekst Kruskala-Wallisa w porównaniach wielokrotnych**

Statystyka testu Kruskala-Wallisa przedstawiona jest w postaci formy kwadratowej z użyciem statystyki Manna-Whitneya lub miar konkordacji τ au Kendall'a.

Na bazie porównań wielokrotnych rozważamy przechodniość i nieprzechodniość efektów zbiegów w jednowymiarowej analizie wariancji.