Best, D. J., Brockhoff, P. B., Rayner, J. C. W., "Tests for balanced incomplete block ranked data with ties ". Originally published in Statistica Neerlandica Vol. 60, Issue 1, p. 3-11 (2006).

Available from: http://dx.doi.org/10.1111/j.1467-9574.2006.00305.x

Tests for balanced incomplete block ranked data with ties

D.J. BEST
School of Mathematics and Applied Statistics, The University of Wollongong, Northfields Avenue, Wollongong, NSW 2522, Australia;
Email: jbest@uow.edu.au

P. B. BROCKHOFF
Informatics and Mathematical Modelling, The Technical University of Denmark, Richard Petersens Plads, Build. 321, DK-2800 Kgs. Lyngby,Denmark;
Email: pbb@imm.dtu.dk

and J.C.W. RAYNER
School of Mathematics and Applied Statistics, The University of Wollongong, Northfields Avenue, Wollongong, NSW 2522, Australia;
Email: john_rayner@uow.edu.au

We consider balanced incomplete block data when ties occur and propose new statistics for testing (a) differences in mean ranks, (b) differences in distributions of ranks, (c) differences in nonlinear effects of ranks and (d) linear contrasts. A sensory evaluation example where the data are ranks is given.

Key words and Phrases: components, Durbin rank test, Nonparametrics, Schach rank test.

1 Introduction

Preference ranking to compare products is well known, and statistical procedures are commonly used to verify that apparent differences between product rankings are due to other than chance effects. Sometimes in taste-testing experiments there are too many products for one judge or consumer to reliably compare at one sitting. This loss of reliability is often associated with sensory fatigue. In such cases balanced incomplete block designs can be employed whereby each judge or consumer tastes only some of the products. The nonparametric analysis described here could, of course, be applied to any set of balanced incomplete block data.

Let N_{ij} be the count in the (i, j)th cell of the $t \times k$ table of counts based on the ranks of t products ranked k at a time by b consumers or judges. The number b is chosen so that if $k < t$ then each product is equally replicated r times, where

$r = bk/t$. This N_{ij} is the number of times that product i receives rank j. If there are no product differences the expectation of N_{ij} is r/k. If X^2 is the usual $\sum(\text{observed} - \text{expected})^2/\text{expected}$ Pearson chi-squared statistic for testing homogeneity of the distributions in a $t \times k$ table of counts, then SCHACH (1979) showed that asymptotically

$$A = \{(t - 1)/t\}X^2$$

has the $\chi^2_{(k-1)(t-1)}$ distribution. The statistics A and X^2 are omnibus statistics.

For data with no ties define the average or linear effect for the ith product, M_i say, as

$$M_i = \sqrt{\left(\frac{t-1}{rt}\right)}\sum_{j=1}^{k} N_{ij} g_1(j)$$

and the quadratic effect for the ith product by

$$V_i = \sqrt{\left(\frac{t-1}{rt}\right)}\sum_{j=1}^{k} N_{ij} g_2(j)$$

in which

$$g_1(j) = \sqrt{\frac{12}{(k^2-1)}}\left\{j - \frac{k+1}{2}\right\}$$

and

$$g_2(j) = \sqrt{\frac{180}{(k^2-1)(k^2-4)}}\left\{\left(j - \frac{k+1}{2}\right)^2 - \frac{k^2-1}{12}\right\}.$$

As M_i is defined in terms of a linear polynomial $g_1(j)$ we say it is a linear effect. Similarly we say V_i is a quadratic effect as $g_2(j)$ is a quadratic polynomial. The definition of M_i involves a difference between the sample mean rank for product i and its expected value assuming a uniform spread of ranks. Similarly V_i involves a difference between the sample variance of the ranks and its expected value assuming a uniform spread of ranks. The values m_i of M_i separate the products according to mean rankings. A high negative value v_i of V_i implies that the rankings are clumped around the middle rankings, whereas a large positive v_i implies that the ranks are at one end or are in two clumps around both high and low rankings. Two clumps indicate either a lack of consensus - market segmentation - or else non-uniform product. One clump indicates consensus.

Clearly to define M_i we need $k > 1$ and to define V_i we need $k > 2$. The statistic $\sum_{i=1}^{t} M_i^2$ is the well-known Durbin rank test that looks for average rank differences between products. The statistic $\sum_{i=1}^{t} V_i^2$ was introduced in RAYNER and BEST (2001, p.123) and assesses dispersion differences between products.

The above statistics allow analysis of ranked data from balanced incomplete block designs when none of the ranks are tied - the forced choice ranking situation. In this note we suggest statistics that allow for tied rankings. In the next section we present an example with tied rankings and give some tables of counts to assist with the analysis. This example is carried through the remainder of the paper. In section 3 we look at linear effects for our example data and in section 4 we extend the statistic *A* to cope with tied ranks. Section 5 considers partitioning the adjusted Durbin statistic. The Appendix shows how to partition the extended *A* statistic so that quadratic and higher order effects can be examined. The statistics we propose are based on those for the randomized complete blocks design given in BROCKHOFF et al. (2004). See further comments in Appendix (ii).

An alternative approach outlined by ALVO and CABILIO (1998) may give a statistic analogous to *A*. We will not discuss this alternative approach here as our purpose is just to present and illustrate our new test statistics. Earlier work on the randomized complete block design was presented by Anderson (1959) and Kannemann (1976).

2 Canned fruits example

Suppose we have a taste-test involving t = 9 brands of canned fruit ranked k = 3 at a time by b = 12 judges, so that each brand is replicated r = 4 times. Acceptability rankings were obtained and are given in Table 1.

Table 1. Acceptability rankings for nine brands of canned fruit

	Brand								
Judge	A	B	C	D	E	F	G	H	I
1	2.5	1	2.5	-	-	-	-	-	-
2	-	-	-	2	1	3	-	-	-
3	-	-	-	-	-	-	3	2	1
4	2.5	-	-	1	-	-	2.5	-	-
5	-	2	-	-	1	-	-	3	-
6	-	-	3	-	-	2	-	-	1
7	3	-	-	-	2	-	-	-	1
8	-	-	3	1	-	-	-	2	-
9	-	1	-	-	-	2	3	-	-
10	-	-	2	-	1	-	3	-	-
11	1.5	-	-	-	-	3	-	1.5	-
12	-	2	-	2	-	-	-	-	2

From Table 1 we can form a summary products (brands) by ranks matrix. For each judge we assign tied ranks by recording $1/m$ for rankings involving an m-way tie. Thus, for example, judge one gives each of brands A and C half a rank of 2 and half a rank of 3 while brand B is given a rank of 1. Table 2 shows the brands by ranks matrix for the first judge only, while Table 3 is formed by summing over all judges, so the counts in Table 3 are the N_{ij} values for the canned fruit data.

Table 2. Brands by ranks matrix of counts for judge 1

Brand	Rank		
	1	2	3
A	0	$\frac{1}{2}$	$\frac{1}{2}$
B	0	1	0
C	0	$\frac{1}{2}$	$\frac{1}{2}$

Table 3. Brands by ranks matrix of counts

Brand	Rank		
	1	2	3
A	$\frac{1}{2}$	$1\frac{1}{2}$	2
B	$2\frac{1}{3}$	$1\frac{1}{3}$	$\frac{1}{3}$
C	0	$1\frac{1}{2}$	$2\frac{1}{2}$
D	$2\frac{1}{3}$	$1\frac{1}{3}$	$\frac{1}{3}$
E	3	1	0
F	0	2	2
G	0	$\frac{1}{2}$	$3\frac{1}{2}$
H	$\frac{1}{2}$	$2\frac{1}{2}$	1
I	$3\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

3 Linear effects

Whether there are ties or not, Table 3 is not the usual type of contingency table as the sum of the row counts is always r. Further, in Table 1 note that the sum of the ranks is always $k(k+1)/2$. These restrictions imply that it is not valid to check the homogeneity of the nine distributions of ranks given in Table 2 by calculating Pearson's X^2. For $k = 3$, $g_1(j) = 1.2247\ (j - 2)$ for $j = 1, 2, 3$ but when there are ties the linear effects, M_i, need adjustment by a factor, a_1 say, which depends on $g_1(j)$ and the ties structure. For when there are ties $\sum_{i=1}^{t} M_i^2$ no longer has a χ^2 distribution. However for large k

$$S = \left(\sum_{i=1}^{t} M_i^2\right) / a_1$$

will have an approximate chi-squared distribution with t - 1 degrees of freedom under the null hypothesis of no treatment effect. Specifically the factor is

$$a_1 = \mathbf{g}_1^{\mathbf{T}}\mathbf{U}\mathbf{g}_1/(rt),$$

where $\mathbf{g}_1^{\mathbf{T}} = (g_1(1), g_1(2), \ldots, g_1(k))$ and the (d, w)th element of $\mathbf{U} = (U_{ij})$ counts the times rank d and rank w are tied. If, for any judge, the ranks $d, \ldots, d + m - 1$ are tied, then $1/m$ is added to each of the m^2 cells in the corresponding submatrix of $\mathbf{U}$, that is, to the elements U_{ij} for $i, j = d, \ldots, d + m - 1$. Clearly the matrix, formed by summing over all judges, is symmetric. For the brand acceptability data Table 4 gives $\mathbf{U}$. To confirm this note that Judge 1, for example, contributes to $\mathbf{U}$ with 0.5 for the (2, 2)th, (2, 3)th, (3, 3)th and (3, 2)th cells and 1 to the (1, 1)th cell. See Table 5. If there are no ties $a_1 = 1$. Here $a_1 = 0.854167$.

Table 4. Ranks by ranks matrix of counts

	Rank		
Rank	1	2	3
1	$10\frac{5}{6}$	$\frac{5}{6}$	$\frac{1}{3}$
2	$\frac{5}{6}$	$9\frac{5}{6}$	$1\frac{1}{3}$
3	$\frac{1}{3}$	$1\frac{1}{3}$	$10\frac{1}{3}$

Table 5. Ranks by ranks matrix of counts for judge 1

	Rank		
Rank	1	2	3
1	1	0	0
2	0	$\frac{1}{2}$	$\frac{1}{2}$
3	0	$\frac{1}{2}$	$\frac{1}{2}$

Using the Table 4 counts we find the adjusted effects $M_i/\sqrt{a_1}$ for $i = 1, \ldots, t$ shown in Table 6. These adjusted effects can then be used to calculate our new Durbin statistic adjusted for ties, $S = \sum_{i=1}^{t} M_i^2 / a_1 = 19.9$ on 8 degrees of freedom giving a p-value close to 0.01 using either the chi-squared distribution or a permutation test.

Table 6. Ordered linear adjusted effects and multiple comparisons

Brand (i)	Linear Effects ($M_i/\sqrt{a_1}$)	Multiple Comparisons
E (5)	-1.8741[a]	a
I (9)	-1.8741[a]	a
B (2)	-1.2494[ab]	ab
D (4)	-1.2494[ab]	ab
H (8)	0.3124[bc]	bc
A (1)	0.9370[bc]	bc
F (6)	1.2494[bc]	bc
C (3)	1.5617[c]	c
G (7)	2.1863[c]	c

Multiple comparisons between the $M_i/\sqrt{a_1}$ can be given using the fact that $(M_i - M_j)/\sqrt{a_1}$ are approximately $N(0, 2)$ when $i \neq j$. This approximation may be justified using an approach similar to that in BROCKHOFF et al. (2004). Table 6 also gives least significant difference multiple comparisons for the present data set. The same superscript indicates the brand linear effects are not significantly different at the 5% level. There are thus three groups judged to be similar: first E, I, B and D, second B, D H, A and F and finally H, A, F, C and G.

4 Schach's statistic for tied ranks

The statistics given so far assess linear effects when there are tied ranks. It is also useful to have an omnibus statistic that potentially detects any sort of differences in the response distributions of the products being compared. For untied ranks SCHACH (1979) gives such an omnibus homogeneity statistic. We now give a generalization of Schach's statistic that allows for tied ranks. For $i = 1, \ldots, t$, let $\mathbf{z_i}$ be vectors of k - 1 elements given by

$$\mathbf{z_i} = ((N_{i1} - r/k)\sqrt{k/r},\ (N_{i2} - r/k)\sqrt{k/r},\ \ldots,\ (N_{i(k-1)} - r/k)\sqrt{k/r})^T.$$

Further, take $\mathbf{R}^* = (\mathbf{U}/b - \mathbf{1}/k)$ where $\mathbf{U}$ is the matrix of counts defined above and $\mathbf{1}$ is the matrix with every element equal to unity. If the data are untied $\mathbf{U}/b$ is the identity matrix. If all possible ranks are untied at least once, a statistic, which reduces to Schach's statistic when there are no ties, is

$$A = \left(\frac{t-1}{t}\right) \sum_{i=1}^{t} \mathbf{z_i^T} \mathbf{R}^{-1} \mathbf{z_i}$$

where $\mathbf{R}$ is $\mathbf{R}^*$ with the last row and column deleted.

For the canned fruit brands acceptability data we have

$\mathbf{z}_1$ = (-0.7217, 0.1444), $\mathbf{z}_2$ = (0.8660, 0.0000), $\mathbf{z}_3$ = (-1.1547, 0.1444)
$\mathbf{z}_4$ = (0.8660, 0.0000), $\mathbf{z}_5$ = (1.4434, -0.2887), $\mathbf{z}_6$ = (-1.1547, 0.5774)
$\mathbf{z}_7$ = (-1.1547, -0.7217), $\mathbf{z}_8$ = (-0.7217, 1.0104), $\mathbf{z}_9$ = (1.7321, -0.8660)

$$\mathbf{R} = \begin{pmatrix} 0.5694 & -0.2639 \\ -0.2639 & 0.4861 \end{pmatrix} \text{ and } \mathbf{R}^{-1} = \begin{pmatrix} 2.3464 & 1.2738 \\ 1.2738 & 2.7486 \end{pmatrix}.$$

We find A = 24.74. By an argument similar to that in BROCKHOFF et al. (2004), A has an approximate chi-squared distribution with (t - 1)(k - 1) = 16 degrees of freedom; see Appendix (ii). We find A has an approximate p-value of 0.07. This p-value is close to the permutation test p-value of 0.05. Durbin's statistic adjusted for ties accounts for most of A (19.9 on 8 degrees of freedom) and so we would conclude that only linear effects are important for this data set.

For tied data we need a new second order polynomial if we wish to have a dispersion test - see Appendix (i).

The calculations outlined in Appendix (i) may not always be needed. A fairly complete analysis may involve calculation of just *A*, Durbin's statistic adjusted for ties and the difference of these two statistics which we *suggest* has an approximate chi-squared distribution with (t - 1)(k - 2) degrees of freedom. This difference statistic indicates whether there are nonlinear effects which, in a sensory evaluation application, could be caused by market segmentation or nonuniform product. Such effects do not appear evident for the canned fruit data.

Although it is not the case here, it can happen that the components of *A* can be significant when *A* is not. A significant value of Durbin's statistic does not always mean *A* will be significant as the omnibus *A* statistic may 'dilute' the effect of the specific component statistics. Thus it is important to look at the components of *A* as well as *A* itself. An omnibus test like *A* has some power against many alternatives but does not always have good power against specific alternatives of interest such as whether the mean ranks differ.

Executable code for a PC which calculates the generalized *A* statistic, as well as the Durbin statistic for ties and permutation test p-values, is available from the first author. If $k = t$ then *A* becomes the generalized Anderson statistic discussed in BROCKHOFF et al. (2004).

5 Partitioning Durbin's statistic

If we want to know, for the Table 1 data, whether the average rank for brand B differs from the average of the average ranks for brand A and C we can form the linear contrast, *L* say, given by

$$1.5a_1 L = \left\{M_2 - (M_1 + M_3)/2\right\}^2.$$

For the Table 1 data, $L = 4.9$ and $S - L = 15.0$ with p-values, based on the approximating chi-squared distributions (with 1 and (t - 2) = 7 degrees of freedom respectively), both less than 0.05. The constant 1.5 is derived, as is usual with linear contrasts, by calculating the sum of squares of the coefficients in the contrast, so that in this case $1.5 = 1^2 + 2(0.5)^2$.

Linear contrasts could also be used to partition the dispersion statistic or other linear contrasts, as appropriate, could be examined.

6 Conclusion

In this note we have considered balanced incomplete block data and proposed new statistics for testing (a) differences in mean ranks when there are ties, (b) differences in distributions of ranks with ties, (c) differences in nonlinear effects of ranks with ties and (d) linear contrasts. We gave a sensory evaluation example where the data were ranks. In some applications the data may be obtained as category rating data or as continuous line scale data and it may be appropriate to rank such data. This might be because, for example, the categories are not equispaced and so assigning scores 1, 2, 3, … is not valid. Alternatively, a consumer who gives continuous line scale scores of 40 and 80 may not really mean that one product had twice the acceptability of the other. With categorical and continuous line scale data perhaps it is worth doing both the normal parametric analysis and the nonparametric ranking analysis we have given here.

Appendix

(i) If we require a dispersion test and the data are tied then we need to redefine $\mathbf{g}_2 = (g_2(j))$. Initially take $g_2(j) = (j^2 + c_1 j + c_0)$ for $j = 1, \ldots, k$ where the constants are to be determined. The orthogonality constraints require $\sum_{j=1}^{t} g_2(j) = 0$ and $\mathbf{g}_1^{\mathbf{T}}\mathbf{U}\mathbf{g}_2 = 0$. This allows us to solve two linear equations for the two unknown constants. We also require a normalizing constant, say E, such that with $g_2(j) = E^{0.5}(j^2 + c_1 j + c_0)$, $\mathbf{g}_2^{\mathbf{T}}(\mathbf{U}/b)\mathbf{g}_2 = k$. Similarly, if $k > 3$, we can redefine $\mathbf{g}_3$, $\mathbf{g}_4$, and so on. BROCKHOFF et al. (2004) discusses the approach in this Appendix in greater detail.

(ii) In the text above we state that certain statistics have approximate normal or chi-squared distributions. This can be verified by use of permutation tests or by easy adjustments to the theory given in BROCKHOFF et al. (2004). If the reader does try these adjustments it is worth noting that for the balanced incomplete block case

$$E(N_{ij}) = r/k$$

and

$$\begin{aligned}
\operatorname{cov}(N_{jl}, N_{j'l'}) &= \frac{\mathbf{U}_{ll'}^{(jj')}}{k} - \frac{b}{kt} \text{ for } j = j' \text{ and } l, l' \text{ tied together} \\
&= \frac{-b}{kt} \text{ for } j = j' \text{ and } l, l' \text{ not tied together} \\
&= \frac{b}{kt(t-1)} - \frac{\mathbf{U}_{ll'}^{(jj')}}{k(k-1)} \text{ for } j \neq j' \text{ and } l, l' \text{ tied together} \\
&= \frac{b}{kt(t-1)} \text{ for } j \neq j' \text{ and } l, l' \text{ not tied together,}
\end{aligned}$$

where $\mathbf{U}^{(jj')}$ for $j = j'$ equals the rank-by-rank matrix of counts based on the r blocks in which treatment $j = j'$ was presented and $\mathbf{U}^{(jj')}$ for $j \neq j'$ equals the rank-by-rank matrix of counts based on the $\lambda = bk(k - 1)/\{t(t - 1)\}$ blocks in which treatment j and j' were presented together. We have suggested following BROCKHOFF et al. (2004) and using a rank-by-rank matrix $\mathbf{U}$ based on all b blocks. This would only be strictly valid if the average tie-structures are the same across all such subsets of blocks

$$\frac{\mathbf{U}^{(jj)}}{r} = \frac{\mathbf{U}^{(jj')}}{\lambda} = \frac{\mathbf{U}}{b}.$$

However, whenever this holds asymptotically,

$$\frac{\mathbf{U}^{(jj)}}{r} - \frac{\mathbf{U}}{b} \to 0 \text{ and } \frac{\mathbf{U}^{(jj')}}{\lambda} - \frac{\mathbf{U}}{b} \to 0,$$

and the approach is still valid.

References

ALVO, M. and P. CABILIO (1998), Application of Hamming distance to the analysis of block designs, in B. Szyszkowicz (ed.), *Asymptotic Methods in Probability and Statistics*, Elsevier Science, Amsterdam, 787-800,.

ANDERSON, R.L. (1959), Use of contingency tables in the analysis of consumer preference studies, *Biometrics* **15**, 582-590.

BROCKHOFF, P.B., D.J. BEST, and J.C.W. RAYNER, (2004), Partitioning Anderson’s statistic with ties, *Journal of Statistical Planning and Inference* **121**, 93-111.

KANNEMANN, K. (1976). An incidence test for *k* related samples, *Biometrische Zeitschrift* **18**, 3-11.

RAYNER, J.C.W., and D.J. BEST (2001). *A contingency table approach to nonparametric testing*, Chapman and Hall/CRC, Boca Raton.

SCHACH, S. (1979), An alternative to the Friedman test with certain optimality properties, *Annals of Statistics* **7**, 537-550.