Metrika (1999) 50: 37–54

Metrika

The analysis of ranked data in blocked factorial experiments

Philip Prescott[1], **Rahim Shahlaee**[2]

[1] Department of Mathematics, University of Southampton, Southampton, SO17 1BJ, UK (e-mail: pp@soton.ac.uk)
[2] Department of Statistics, Faculty of Mathematical Sciences, Shaheed Beheshty University, Tehran, Evin 19832, Iran (e-mail: a-shahlaee@cc.sbu.ac.ir)

Received: February 1999

Abstract. A non-parametric method for the analysis of blocked factorial experiments, based on ranking within blocks, is proposed and shown to be equivalent to partitioning Friedman's test statistic into a set of contrasts reflecting polynomial components of the main effects and interaction. A slightly modified version of the procedure is suggested to partially overcome the problem of loss of power to detect one component when the model includes other components. This alternative procedure is shown to be equivalent to applying a standard normal theory analysis of variance to the ranks. The null distributions and power comparisons are investigated using simulation methods, and it is shown that the non-parametric methods are almost as powerful as the analysis of variance.

Key words: Analysis of variance, blocking in factorial experiments, Friedman's test, orthogonal contrasts, ranking

1 Introduction

Factorial experiments in plant research are often arranged in blocks such that a full, or fractional, replicate is contained within each block. Recently, a problem arose in a cereal crop context, which involved measuring the amount of disease that developed on the leaves of plants given different amounts of nitrate fertilizer and subjected to different concentrations of carbon dioxide. The experiment was a 6×2 factorial design arranged in blocks, to take account of the position in the greenhouse. However, the blocks were of sufficient size to allow one plant in each block to be subjected to each of the 12 fertilizer-carbon dioxide treatment combinations. One of the major problems with experiments of this kind is that it is difficult and time consuming to measure the absolute amount of disease on each plant. However, ranking the plants within each block is simple, but this requires an appropriate rank-based analysis capable of assessing the main effects of the factors and their interaction. The method should also be able to investigate the linear and quadratic components of the main effects and the interactions using only the ranks within the blocks.

The analyses of one-way and two-way layouts using the Kruskal-Wallis and Friedman test procedures are well established, but extensions to these tests to other situations are limited. For the one-way layout involving a single factor at several different levels, the general alternative hypothesis of at least one location parameter being different from the rest might not be of primary interest. A researcher might be more interested in detecting a linear trend in the location parameters over the levels of the factor. Terpstra (1953) and Jonckheere (1954) independently proposed a test for the ordered alternative based on the pairwise Mann-Whitney-Wilcoxon statistics. A weighted linear combination of such pairwise statistics was considered by Tyron and Hettmansperger (1974), see also Barlow et al. (1972) for a review of inferences under order restrictions.

For the two-way layout, an analogue of the test proposed by Jonckheere and Terpstra has been studied by Skillings and Wolfe (1978), for the problem of detecting a trend in the location parameters of one factor, with a blocking factor taken into account. This test is again based on pairwise Mann-Whitney-Wilcoxon statistics, but in this case applied within each block. An alternative, studied by Hollander (1967) and Puri and Sen (1968), relies on the sum of a series of Wilcoxon signed rank tests computed on the ith and jth paired samples.

The application of ranking methods to data obtained from completely randomised factorial designs was considered by Scheirer et al. (1976). They presented a method based on the Kruskal-Wallis test applied to a ranking of the entire data set consisting of an equal number of observations for each treatment combination. Simulated null distributions were produced for test statistics defined to reflect main effects and interactions of the factors. Iman (1974) and Conover and Iman (1976) used empirical methods to investigated the use of the rank transform to detect main effects and interactions in the two-factor case; see also Lemmer (1980). Scheirer et al. (1976) partitioned the Kruskal-Wallis test statistic into components for main effects and interactions. Further investigation of the rank transform, see for example, Brummer and Neumann (1986), Blair et al. (1987), Sawilowsky et al. (1989) and Akritas (1990), suggested that, because of the non-linear nature of the rank transform, spurious indications of significant effects could occur when other effects are presence. An alternative ranking method for factorial experiments, in which observations are ranked for the levels of one factor within the levels of the other factor, was considered by Shirley (1987). Details of other procedures using ordered categorical data from factorial experiments, may be found in Thomas and Kiwanga (1993).

In our practical problem, we do not have numerical data, simply the ranks within blocks, so it is not possible to obtain the overall rank transform or to rank the data over the different levels of the factors. We do, however, have a series of replications of the factorial experiment over the blocks and a set of ranks for each of these replications. In the following section, we consider a simple extension of the Friedman test statistic, which involves partitioning it into components for trends in the main effects and interactions and we investigate the properties of these components. Because of certain limitations of this extended Friedman test, a modified form of the test procedure is considered, and a power comparison is used to indicate its performance relative to the analysis of variance.

2 The underlying model and orthogonal contrasts

Suppose that we have a two-factor experiment with factors U and V having I and J levels respectively. We restrict attention to the case where the levels of the two factors are equally spaced on a linear or some transformed scale, but the basic ideas may be extended to the more general case of more than two factors and unequally spaced levels. The experiment is conducted in a randomised block design with M blocks, each block containing a full replicate of IJ treatment combinations. The response variable is y_{ijm} for the ith level of factor U, the jth level of factor V within the mth block, where $i = 1, \dots, I$, $j = 1, \dots, J$ and $m = 1, \dots, M$. If y_{ijm} were a continuous response variable, then a suitable model would be

$$y_{ijm} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_m + \varepsilon_{ijm} \tag{1}$$

where μ is a mean level, α_i, $i = 1, \dots, I$, is the fixed effect of level i of factor U, β_j, $j = 1, \dots, J$, is the fixed effect of level j of factor V, $(\alpha\beta)_{ij}$, $i = 1, \dots, I$ and $j = 1, \dots, J$, is the interaction between level i of factor U and level j of factor V, γ_m, $m = 1, \dots, M$, is the fixed or random effect of block m and ε_{ijm} is an error, independent of any other error, with an assumed distribution, typically $N(0, \sigma^2)$.

An appropriate analysis for this situation is an analysis of variance (ANOVA), in which the main effects and interaction are examined using F-tests with $I-1$, $J-1$ and $(I-1)(J-1)$ degrees of freedom respectively. These sums of squares could be divided into orthogonal components to investigate any linear or quadratic trends in the main effects or any linear by linear trends, etc in the interaction. The significance of these orthogonal components is assessed by comparison with an error component from which the block component has been removed. The ANOVA procedure is based on a set of orthogonal contrasts of the form

$$T = \sum_{k=1}^{K} c_k y_{\bullet k} \tag{2}$$

where c_k are suitable coefficients used to identify linear, quadratic etc trends, $y_{\bullet k}$ is the sum of the responses for the kth treatment combination over the M blocks and $k = 1, \dots, K$ with $K = IJ$. For this T to represent a normalized contrast, we must also have that $\sum_{k=1}^{K} c_k = 0$ and $\sum_{k=1}^{K} c_k^2 = 1$. (Note that the coefficients should be scaled by an additional factor $\sqrt{M}$ since they are applied to sums over M blocks. However, it is more convenient to define the coefficients using $\sum_{k=1}^{K} c_k^2 = 1$ as these may be found in standard tables of orthogonal polynomial coefficients.)

Table 1. Ranked data for the treatment combinations in a randomised block design

	Treatment combinations				
Blocks	1	2	3	. . .	K
1	R_{11}	R_{12}	R_{13}	. . .	R_{1K}
2	R_{21}	R_{22}	R_{23}	. . .	R_{2K}
3	R_{31}	R_{32}	R_{33}	. . .	R_{3K}
.	.	.	.	. . .	.
.	.	.	.	. . .	.
.	.	.	.	. . .	.
M	R_{M1}	R_{M2}	R_{M3}	. . .	R_{MK}
Totals	$R_{\bullet 1}$	$R_{\bullet 2}$	$R_{\bullet 3}$	. . .	$R_{\bullet K}$

When individual responses are not available, or are difficult to obtain, and ranks are used instead, it is still of interest to investigate trends in the main effects and interactions of the factors. Since the ranking is carried out within blocks, the ranks range from 1 to K, and $y_{\bullet k}$, for $k = 1, \ldots, K$, in the definition of the contrast T can be replaced by the sum of the ranks for each treatment combination over the M blocks. The structure of the two-way layout with ranks replacing the responses is illustrated in Table 1.

In the table, R_{mk} is the rank of the kth treatment combination in the mth block so that

$$\sum_{k=1}^{K} R_{mk} = \frac{K(K+1)}{2} \quad \text{for } m = 1, \ldots, M \tag{3}$$

and $R_{\bullet k} = \sum_{m=1}^{M} R_{mk}$, $k = 1, \ldots, K$, is the sum of the ranks for the kth treatment combination. A contrast T may now be defined as $T = \sum_{k=1}^{K} c_k R_{\bullet k}$, where the coefficients c_k, $k = 1, \ldots, K$, as before, are suitably chosen to reflect the specific linear, quadratic and other components of the main effects or the linear by linear, linear by quadratic and other components of the interaction. Since the main effects and interaction have a total of $(I-1)+(J-1)+(I-1)(J-1) = IJ-1 = K-1$ degrees of freedom, there will be $K-1$ contrasts into which the main effects and interaction can be partitioned. This complete set of contrasts is defined as

$$T_l = \sum_{k=1}^{K} c_{lk} R_{\bullet k} \quad \text{for } l = 1, \ldots, K-1 \tag{4}$$

where c_{lk}, $l = 1, \ldots, K-1$, and $k = 1, \ldots, K$, are such that $\sum_{k=1}^{K} c_{lk} = 0$, $\sum_{k=1}^{K} c_{lk}^2 = 1$ and $\sum_{k=1}^{K} c_{l_1 k} c_{l_2 k} = 0$ for $l_1 \neq l_2$ and $l_1, l_2 = 1, \ldots, K-1$.

2.1 Expectations, variances and covariances of the contrasts

Theorem 1. *Under the null hypothesis of no main effects and interaction, ie under H_0 given by*

$$H_0 : \alpha_i = 0, \beta_j = 0 \quad \textit{and} \quad (\alpha\beta)_{ij} = 0 \quad \textit{for all } i, j,$$

the mean and variance of the contrasts are given by $E(T_l) = 0$ *and* $\mathrm{Var}(T_l) = MK(K+1)/12$, *for* $l = 1, \dots, K-1$, *and the covariance between any pair of contrasts is* $\mathrm{Cov}(T_{l_1}, T_{l_2}) = 0$.

Proof. Let $\mathbf{R}' = (R_{\bullet 1}, \dots, R_{\bullet K})$ be the vector of rank sums and let $\mathbf{c}_l' = \{c_{lk}\}$ be the vector of contrast coefficients c_{lk} for $l = 1, \dots, K-1$ and $k = 1, \dots, K$, and let $\mathbf{1}_K$ be the vector of K 1's. Then $T_l = \mathbf{c}_l'\mathbf{R}$, where $E(\mathbf{R}) = MK(K+1)/2$ and $V(\mathbf{R}) = MK(K+1)/12\{\mathbf{I}_K - \mathbf{1}_K\mathbf{1}_K'/K\}$. Thus, $E(T_l) = \mathbf{c}_l'E(\mathbf{R}) = MK(K+1)/2\mathbf{c}_l'\mathbf{1}_k = 0$, and the covariance of any two contrasts is $\mathrm{Cov}(T_{l_1}, T_{l_2}) = MK(K+1)/12\mathbf{c}_{l_1}'\{\mathbf{I}_K - \mathbf{1}_K\mathbf{1}_K'/K\}\mathbf{c}_{l_2}$ which equals zero if $l_1 \neq l_2$, but equals $MK(K+1)/12$ if $l_1 = l_2$.

Note that even though the ranks within a block are not independent of each other, the orthogonality condition implies that the contrasts T_l for $l = 1, \dots, K-1$ involving the rank sums are uncorrelated.

2.2 Relationship with Friedman's statistic

The Friedman statistic for testing for differences between the location parameters of the different treatment combinations based on the rankings shown in Table 1 is given by

$$F = \frac{12}{MK(K+1)} \sum_{k=1}^{K} R_{\bullet k}^2 - 3M(K+1). \tag{5}$$

Values of F larger than the upper α percentile of the chi-squared distribution with $K-1$ degrees of freedom would lead to rejection of the null hypothesis of equality of the location parameters.

Theorem 2. *For the set of contrasts defined in equation (4), Friedman's test statistic may be expressed as*

$$F = \frac{12}{MK(K+1)} \sum_{l=1}^{K-1} T_l^2. \tag{6}$$

Proof. Let $T_K = \sum_{k=1}^{K} c_{Kk}R_{\bullet k}$ with $c_{Kk} = 1/\sqrt{K}$ for $k = 1, \dots, K$, then $\sum_{k=1}^{K} c_{Kk}^2 = 1$ and $\sum_{k=1}^{K} c_{Kk}c_{lk} = 0$ for $l = 1, \dots, K-1$. Let $\mathbf{T}' = (T_1, \dots, T_K)$ be the augmented vector of contrasts and $\mathbf{C} = \{c_{lk}\}$ be the $K \times K$ matrix of contrast coefficients c_{lk} for $l = 1, \ldots, K$ and $k = 1, \ldots, K$. Then $\mathbf{T} = \mathbf{CR}$ and $\mathbf{C}'\mathbf{C} = \mathbf{I}$, so that $\sum_{l=1}^{K} T_l^2 = \mathbf{T}'\mathbf{T} = \mathbf{R}'\mathbf{C}'\mathbf{C}\mathbf{R} = \mathbf{R}'\mathbf{R} = \sum_{k=1}^{K} R_{\bullet k}^2$.

Now $T_K = \sum_{k=1}^{K} R_{\bullet k}/\sqrt{K} = \dfrac{MK(K+1)}{2\sqrt{K}}$ so that $T_K^2 = \dfrac{M^2K(K+1)^2}{4}$ and therefore $\dfrac{12}{MK(K+1)} T_K^2 = 3M(K+1)$.

But $F = \dfrac{12}{MK(K+1)} \sum_{k=1}^{K} R_{\bullet k}^2 - 3M(K+1) = \dfrac{12}{MK(K+1)} \sum_{l=1}^{K} T_l^2 - 3M(K+1) = \dfrac{12}{MK(K+1)} \sum_{l=1}^{K-1} T_l^2$, which completes the proof.

Friedman's test statistic is asymptotically χ_{K-1}^2 as $M \to \infty$ and is represented as the sum of squares of a set of uncorrelated contrasts $T_l/\{\mathrm{Var}(T_l)\}^{1/2}$ each of which is asymptotically $N(0,1)$. Specific contrasts may be tested either using the normal distribution or by referring $\dfrac{12}{MK(K+1)} T_l^2$ to the percentage points of the chi-squared distribution with 1 degree of freedom for $l = 1, \ldots, K-1$.

3 Exact null distributions for small designs

In this section we examine the exact distributions of the test statistics $T_l/\{\mathrm{Var}(T_l)\}^{1/2}$, for $l = 1, \ldots, K-1$, for some selected small designs with parameters (I, J, M). Within each of the M blocks, the ranks allocated are $1, 2, \ldots, K = IJ$, assuming there are no ties. The total number of arrangements of these ranks is $(K!)^M$ since there are $K!$ arrangements within each block. For small designs, complete enumeration of the values of specific contrasts may be carried out for all possible configurations of the ranks, to determine the exact distributions of the test statistics under the null hypothesis of no main effects or interactions. To illustrate these results, the exact distributions of the standardised contrast for the linear component of U given by $T_1/\{12/MK(K+1)\}^{1/2}$, with T_1 suitably defined, are shown in Table 2 for the three small designs $(2,2,2)$, $(2,2,3)$ and $(2,3,3)$, (probabilities for negative values are obtained by symmetry about zero). Note that these null distributions are discrete in nature and that their form depends on the number of blocks M.

3.1 Comparison with the normal approximation

The exact distributions for larger designs could be determined in a similar way, although this becomes difficult and tedious as the number of treatment combinations and blocks are increased. It is evident that the number of discrete values taken by the test statistics increases quickly as the designs become larger. In general, each orthogonal polynomial component for the main effects and interaction, within a design, has a different distribution since the coefficients used to define the test statistics are different (although there is some

Table 2. Exact distributions of the linear U component for several small designs

(2, 2, 2)		(2, 2, 3)		(2, 3, 2)	
$T_l/\{\mathrm{Var}\,T_l\}^{1/2}$	Probability	$T_l/\{\mathrm{Var}\,T_l\}^{1/2}$	Probability	$T_l/\{\mathrm{Var}\,T_l\}^{1/2}$	Probability
0	0.2222	0	0.1759	0	0.1200
0.548	0.1667	0.447	0.1528	0.309	0.1125
1.095	0.1389	0.894	0.1250	0.617	0.1000
1.643	0.0556	1.342	0.0741	0.926	0.0825
2.191	0.0278	1.789	0.0417	1.234	0.0600
		2.236	0.0139	1.543	0.0400
		2.683	0.0046	1.852	0.0250
				2.160	0.0125
				2.469	0.0050
				2.777	0.0025

Only the non-negative values are shown since the distributions are symetric.

Table 3. Comparison of exact and normal percentage points for some small designs

		Percentage points	
		97.5%	99.5%
Normal approximation		1.960	2.576
Design	Contrast		
(2, 2, 2)*	Linear U	2.191	–
(2, 2, 3)*	Linear U	1.789	2.236
(2, 3, 2)*	Linear U	1.852	2.469
	Linear V	1.890	2.457
	Quad V	1.964	2.291
	Lin $U \times$ Lin V	1.890	2.457
	Lin $U \times$ Quad V	1.964	2.400

obvious reduction in this when the number of levels of the two factors are the same). It would be a major practical disadvantage to the use of this non-parametric procedure if extensive tables of percentage points were needed for its implementation. It is of interest, therefore, to investigate whether a normal approximation is satisfactory for designs of a practical size. Table 3 shows the percentage points of the distributions of the test statistics for all the components for a range of designs, based on 100,000 simulations of the experiment.

The simulations were checked by comparison with exact results, where available, and these gave virtually identical values. Table 4 shows the 97.5 and 99.5 percentage points of the distributions for each component (where different) for a range of designs up to $(2, 5, 5)$. It is evident from Table 4 that, provided the design is not very small, the normal approximation seems to be quite adequate even when one factor is at only two levels. For practical purposes, we have defined a non-parametric procedure for assessing the components in a factorial experiment arranged in blocks, where the responses are ranked within a block. The test statistics represent a partition of Friedman's test into a set of $K-1$ components which, under the null hypothesis of no main effects and interaction, are uncorrelated with asymptotic chi-squared distributions

Table 4. Comparison of simulated and normal percentage points for several designs

		Percentage points	
		97.5%	99.5%
Normal approximation		1.960	2.576
Design	Contrast		
(2, 3, 3)	Linear U	1.890	2.646
	Linear V	2.006	2.469
	Quad V	1.871	2.405
	Lin $U \times$ Lin V	2.006	2.469
	Lin $U \times$ Quad V	1.960	2.405
(3, 3, 2)	Linear U	1.897	2.424
	Quad U	2.008	2.556
	Lin $U \times$ Lin V	1.936	2.453
	Lin $U \times$ Quad V	1.938	2.460
	Quad $U \times$ Quad V	1.936	2.453
(3, 3, 3)	Linear U	1.980	2.496
	Quad U	1.938	2.534
	Lin $U \times$ Lin V	1.897	2.530
	Lin $U \times$ Quad V	1.947	2.495
	Quad $U \times$ Quad V	1.897	2.530
(2, 5, 5)	Linear U	1.915	2.569
	Linear V	1.949	2.510
	Quad V	1.954	2.596
	Cubic V	1.949	2.543
	Quartic V	1.997	2.559
	Lin $U \times$ Lin V	1.949	2.510
	Lin $U \times$ Quad V	1.954	2.540
	Lin $U \times$ Cubic V	1.949	2.510
	Lin $U \times$ Quartic V	1.960	2.534

each with one degree of freedom. These components may be tested as χ_1^2 variables, or as standard normal variables using $T_l/\{\mathrm{Var}(T_l)\}^{1/2}$.

One problem with this procedure is that the null hypothesis of interest is not that there are no main effects and interactions present, i.e. that $\alpha_i = 0$, for $i = 1, \ldots, I$, *and* $\beta_j = 0$, for $j = 1, \ldots, J$ *and* $(\alpha\beta)_{ij} = 0$, for $i = 1, \ldots, I$, $j = 1, \ldots, J$, but that $\alpha_i = 0$, for $i = 1, \ldots, I$, *or* $\beta_j = 0$, for $j = 1, \ldots, J$ *or* $(\alpha\beta)_{ij} = 0$, for $i = 1, \ldots, I$, $j = 1, \ldots, J$. Specifically, it is desirable to be able to test whether there are significant components of the main effect of one of the factors when the other factor is present in the model, and also whether the interaction is significant when main effects of both factors are present. We shall consider this in the following section where we examine the powers of these rank-based test statistics to detect polynomial contrasts of varying magnitudes in the presence of other contrasts.

4 Examination of the powers of the rank-based test statistics relative to the analysis of variance

The motivation for this non-parametric procedure came from an experimental situation where plants were to be ranked for leaf disease to avoid the difficulties of exact measurement of the response. In such a situation it would not be possible to carry out a detailed conventional analysis of variance since only the ranks would be available. However, in this section we examine the power of the rank-based method (RANK) relative to the analysis of variance (ANOVA) using orthogonal polynomial contrasts given by equations (4) and (2) respectively. We assume that both factors are quantitative at equally spaced levels and that the responses were generated according to one of several models with main effects and interactions of selected magnitudes present. Simulation methods based on 30,000 experiments were used to compare the powers of the RANK and ANOVA tests against a variety of non-null situations. In these power calculations, the RANK statistics were compared with their simulated null percentage points rather than the asymptotic values, so that the comparisons with ANOVA were not affected by any increase in the type I error.

The model for the alternative hypothesis with linear main effects and linear by linear interaction used for the simulated power comparisons was of the form

$$H_{U,V,UV} : y_{ijm} = g_1(i - \bar{u}_i) + g_2(j - \bar{v}_j) + g_3(i - \bar{u}_i)(j - \bar{v}_j) + \varepsilon_{ijm}, \qquad (7)$$

where $\bar{u}_i$ and $\bar{v}_j$ are the mean levels for factor U and V, with levels $i = 1, \ldots, I$ and $j = 1, \ldots, J$ respectively, and ε_{ijm} is $N(0, \sigma^2)$ with σ^2 taken to be 1. The magnitudes of the linear components and interaction were varied by changing the values of the multipliers g_1, g_2 and g_3. Comparisons of the powers for the two procedures, RANK and ANOVA, were carried out by evaluating the various test statistics for each simulated sample obtained using a range of combinations of g_1, g_2 and g_3 values. In order to simplify the presentation of the results of these simulations, we shall present the results of the comparisons for three different models. Firstly, the alternative model contains only a linear component of the main effect of factor U, i.e. $g_1 = g$, and $g_2 = g_3 = 0$ which corresponds to the hypothesis $H_{U,0,0}$. Secondly, when $g_1 = g_2 = g$ and $g_3 = 0$, the alternative model contains only a linear component of U and a linear component of V of the same magnitude, corresponding to the hypothesis $H_{U,V,0}$, and, finally, when $g_1 = g_2 = g_3 = g$, the model has linear components of both factors and the linear by linear component of the interaction present: this corresponds to the hypothesis $H_{U,V,UV}$.

The orthogonality of the contrasts used to test these components with the ANOVA ensures independence of the contrasts under both the null and the alternative models, so that the power of the ANOVA procedure to detect the linear component of factor U (represented by values of $g_1 = g$) is not affected by the presence or absence of other effects (represented by the values of g_2 and g_3). However, the same is not true of the ranking procedure, since the contrasts in the rank sums, although uncorrelated under $H_{0,0,0}$ are not necessarily uncorrelated under $H_{U,0,0}$, $H_{U,V,0}$ or $H_{U,V,UV}$. The power of RANK to detect one component, for example the linear component of U, will be affected by the presence of another component such as the linear effect of V or the linear by linear component of the interaction. This interdependence is related to the fact that, under the alternative models, the variances of the ranks, and therefore the variances of the rank sums, are not the same as under the null model, $H_{0,0,0}$. This feature of ranking methods has been noted elsewhere. For example, Shirley (1987) commented on the over-estimation of the variance of the ranks due to some subsets of the data being "constrained" in the non-null situation. See also Steel (1960), Shorack (1967) and Williams (1986).

Table 5 shows the powers of ANOVA and RANK for the design with both factors at four levels within each of two blocks, i.e. design (4, 4, 2). Powers are given for testing for the presence of the linear component of factor U using models $H_{U,0,0}$, $H_{U,V,0}$ and $H_{U,V,UV}$, and for the linear × linear component of the interaction using models $H_{0,0,UV}$, $H_{U,0,UV}$ and $H_{U,V,UV}$. In this way we may compare the powers of RANK and ANOVA to detect an effect with only that particular effect present and in the presence of other effects.

It may be seen from Table 5, which is typical of many comparisons carried out for other designs, that the power of RANK compares very favourably with that of ANOVA when no other effects are present in the model. The loss of power resulting from the use of ranks is about 0.5 per cent when testing for the linear component of U and about 1 per cent when testing for the linear by linear component of the interaction when no other effect is present in the model. However, the powers of the ranking procedure are considerably lower than those of ANOVA when the model includes other effects. For example, when the alternative model contains both linear components and the linear by linear component of the interaction, the power of the tests based on the rank sums are only about 70% of the powers of the corresponding ANOVA tests. These power comparisons are illustrated in Figures 1a and 1b.

The reason for the loss of power is that, when other effects are present in the model, the standardisation of T_l by its null standard deviation produces a test statistic which is generally reduced in magnitude relative to the corresponding ANOVA test statistic. This null standard deviation, $\{12/MK(K+1)\}^{1/2}$, is an over-estimate of the appropriate standard deviation of T_l when the ranks are constrained as they would be when the model contains additional components.

5 An alternative non-parametric procedure

In order to overcome this problem of loss of power, at least to some extent, we propose a modification to the non-parametric procedure RANK, so that the previously defined T_l^2, $l = 1, \ldots, K-1$, are compared to an estimate of between-block variability instead of to their null variance. Because two or more blocks are used in the designs under consideration, the between-block information can be used as a measure of variability. For each treatment combination, i.e. for each value of $k, k = 1, \ldots, K$, the ranks are given by R_{mk}, $m = 1, \ldots, M$, so that the variability in these ranks may be assessed using

$$\sum_{m=1}^{M} (R_{mk} - R_{\bullet k}/M)^2. \tag{8}$$

The overall between-block variability is given by

$$\sum_{k=1}^{K} \sum_{m=1}^{M} (R_{mk} - R_{\bullet k}/M)^2, \tag{9}$$

which has $(M-1)(K-1)$ degrees of freedom, since the ranks within each

Table 5. Powers of ANOVA and RANK testing at the 5% level for a linear component of factor U (or factor V) and a linear × linear component of the interaction against various alternative models for the design $(4, 4, 2)$

Test for linear component					Test for linear × linear component				
g	ANOVA TEST	RANK TEST			g	ANOVA TEST	RANK TEST		
		$(U, 0, 0)$	$(U, V, 0)$	(U, V, UV)			$(0, 0, UV)$	$(U, 0, UV)$	(U, V, UV)
0	0.047	0.047	0.047	0.047	0	0.051	0.051	0.051	0.051
0.08	0.077	0.073	0.071	0.069	0.08	0.083	0.082	0.080	0.079
0.16	0.153	0.147	0.139	0.130	0.16	0.185	0.179	0.169	0.161
0.24	0.302	0.290	0.265	0.228	0.24	0.357	0.344	0.308	0.276
0.32	0.474	0.453	0.400	0.326	0.32	0.564	0.545	0.476	0.412
0.40	0.666	0.640	0.558	0.433	0.40	0.757	0.737	0.639	0.542
0.48	0.810	0.792	0.692	0.526	0.48	0.885	0.871	0.768	0.645
0.56	0.912	0.900	0.806	0.612	0.56	0.959	0.951	0.862	0.728
0.64	0.965	0.958	0.883	0.682	0.64	0.988	0.985	0.921	0.798
0.72	0.990	0.987	0.938	0.745	0.72	0.998	0.996	0.956	0.844

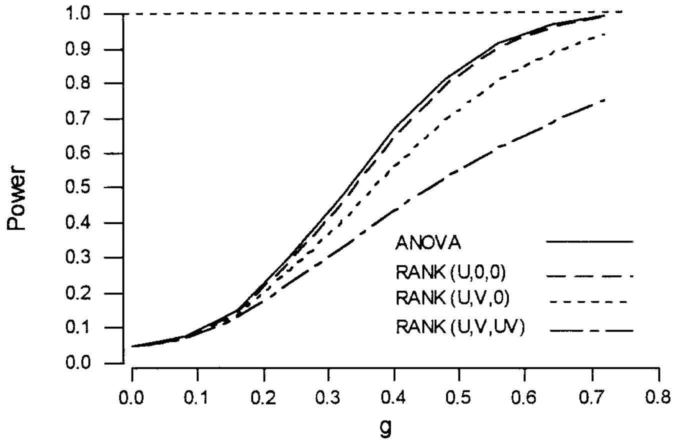

Fig. 1a. Comparison of powers of ANOVA and RANK for testing the linear component of U for the design $(4, 4, 2)$

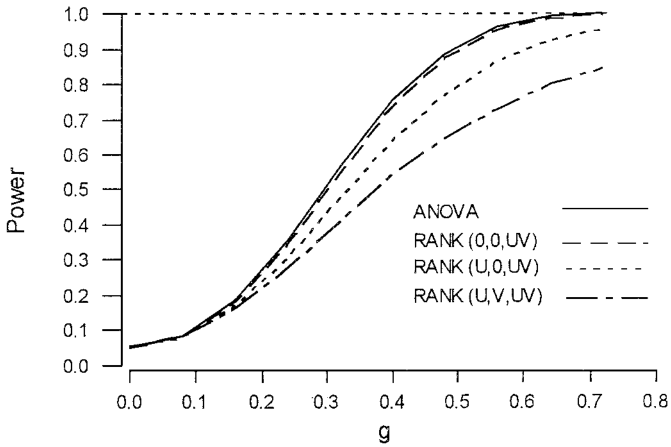

Fig. 1b. Comparison of powers of ANOVA and RANK for testing the linear by linear component of the UV interaction for the design $(4, 4, 2)$

block sum to $K(K+1)/2$. Comparison of the contrasts T_l, $l = 1, \ldots, K-1$, with this "error mean square" leads to the test statistics

$$F_l = \frac{T_l^2/M}{\sum_{k=1}^{K}\sum_{m=1}^{M}(R_{mk} - R_{\bullet k}/M)^2/(M-1)(K-1)}, \tag{10}$$

for $l = 1, \ldots, K-1$. The asymptotic distribution of F_l is an F distribution with 1 and $(M-1)(K-1)$ degrees of freedom. This is equivalent to carrying out a normal theory analysis of variance, in which the responses are the ranks as described in Table 1, with the treatment sum of squares divided into the various linear, quadratic etc components given by T_l^2/M for $l = 1, \ldots, K-1$. In this analysis there is a block component with $(M-1)$ degrees of freedom with a sum of squares equal to zero and an error sum of squares with $(M-1)(K-1)$ degrees of freedom as in Equation (9). Use of Equation (10) is analogous to using the same ANOVA procedure for the blocked factorial experiment as used earlier but with the ranks replacing the original observations. We shall refer to the non-parametric procedure based on testing for the various components of the main effects and interaction using Equation (10) as ANALOGUE.

5.1 The null distribution of the ANALOGUE test

One of the difficulties introduced in suggesting ANALOGUE as an alternative to RANK, is that it is possible for the denominator to equal zero so that F_l is infinite. This occurs with non-negligible probability for some small designs when the rankings within each block are identical. Under the null hypothesis, for the design $(2, 2, 2)$, the probability of the error mean square being zero is $1/4!$ $(=0.0417)$, whereas for $(2, 3, 2)$, this is equal to $1/6!$ $(=0.00139)$ and for $(2, 5, 2)$, it is $1/10!$ $(=2.75 \times 10^{-7})$. Evidently care will be required in the use of the ANALOGUE procedure for very small designs. As with the RANK procedure, a series of simulations was carried out to obtain the null distributions of F_l for a selection of designs. Table 6 gives a summary of the comparisons with the percentage points of the appropriate asymptotic F distributions.

The results indicate that, even with only two blocks, the asymptotic percentage points should be reasonable for practical purposes with all but the smallest designs. For designs such as $(3, 3, 2)$ and $(2, 5, 2)$, the use of the percentage points of $F_{1,8}$ and $F_{1,9}$ will result in slightly larger type I errors than desired, but for $(4, 4, 2)$ and larger designs, it seems that the asymptotic percentage points correspond very well to the simulated values, so that the type I error should be close to the specified levels.

6 Power comparison of ANALOGUE, RANK and ANOVA

To investigate whether the adaptation of the RANK method to ANALOGUE has resulted in an improved power when additional effects are present in the model, further simulations were carried out for a range of designs. Again, as in Section 4, all powers of the various test statistics were obtained by comparison with the corresponding simulated percentage points of their null distributions, so that appropriate type I errors were employed.

From the large number of comparisons made for a range of designs and with models including main effects and interactions of various magnitudes, the results for design $(4, 4, 2)$ are given in Table 7 corresponding to the results shown in Table 5.

The powers shown in Table 5 for RANK and Table 7 for ANALOGUE were based on the same simulated data, so that comparisons can be confidently made both within and between the tables. From Table 7, it is evident

Table 6. Comparison of simulated percentage points and corresponding F percentage points for a selection of designs (I, J, M)

Design	Contrast	95%	99%
(2, 3, 2)	Asymptotic $F_{1,5}$	6.608	16.260
	Linear U	8.167	26.667
	Linear V	8.182	25.312
	Quad V	8.352	23.438
	Lin $U \times$ Lin V	8.000	25.312
(2, 4, 2)	Asymptotic $F_{1,7}$	5.591	12.250
	Linear U	6.034	14.175
	Linear V	6.050	14.787
	Quad V	6.034	14.787
	Cubic V	6.050	14.450
	Lin $U \times$ Lin V	5.973	14.400
(3, 3, 2)	Asymptotic $F_{1,8}$	5.318	11.260
	Linear U	5.633	12.800
	Quad U	5.556	12.522
	Linear V	5.597	12.789
	Quad V	5.628	13.000
	Lin $U \times$ Lin V	5.633	12.600
(2, 5, 2)	Asymptotic $F_{1,9}$	5.117	10.560
	Linear U	5.358	12.166
	Linear V	5.326	11.719
	Quad V	5.403	11.716
	Cubic V	5.409	11.912
	Quartic V	5.369	11.716
	Lin $U \times$ Lin V	5.358	11.683
(4, 4, 2)	Asymptotic $F_{1,15}$	4.543	8.683
	Linear U	4.576	8.883
	Quad U	4.601	9.007
	Linear V	4.474	8.606
	Quad V	4.559	8.744
	Lin $U \times$ Lin V	4.510	8.670

that the powers of ANALOGUE for testing for an effect when only that effect is present compare favourably with ANOVA, although they are not quite as close as the powers of RANK. However, the powers of ANALOGUE for testing for one effect in the presence of other effects are considerably better than those for RANK. For example, the power of ANALOGUE is about 20% higher than RANK for identifying the linear component of factor U when both the linear component of factor V and the linear by linear component of the interaction are present in the model. The adaptation of the non-parametric procedure seems to have overcome the power loss to a considerable extent. The limited results illustrated in Tables 5 and 7 are typical of many other designs investigated. Figures 2a and 2b illustrate the improvement in the powers of ANALOGUE relative to RANK for the design $(4, 4, 2)$ for testing for the linear effect of U and for testing for the linear by linear effect of the interaction when all three components are included in the model.

Table 7. Powers of ANOVA and ANALOGUE testing at the 5% level for a linear component of factor U (or factor V) and a linear $\times$ linear component of the interaction against various alternative models for the design $(4, 4, 2)$

Test for linear component					Test for linear × linear component				
g_1	ANOVA TEST	ANALOGUE TEST			g_3	ANOVA TEST	ANALOGUE TEST		
		$(U, 0, 0)$	$(U, V, 0)$	(U, V, UV)			$(0, 0, UV)$	$(U, 0, UV)$	(U, V, UV)
0	0.047	0.050	0.050	0.050	0	0.051	0.052	0.052	0.052
0.08	0.077	0.077	0.077	0.076	0.08	0.083	0.082	0.082	0.082
0.16	0.153	0.148	0.148	0.146	0.16	0.185	0.175	0.173	0.171
0.24	0.302	0.288	0.284	0.273	0.24	0.357	0.331	0.321	0.311
0.32	0.474	0.446	0.442	0.417	0.32	0.564	0.528	0.509	0.486
0.40	0.666	0.632	0.626	0.575	0.40	0.757	0.714	0.686	0.653
0.48	0.810	0.779	0.771	0.707	0.48	0.885	0.852	0.822	0.779
0.56	0.912	0.887	0.879	0.813	0.56	0.959	0.937	0.912	0.874
0.64	0.965	0.950	0.946	0.887	0.64	0.988	0.978	0.963	0.934
0.72	0.990	0.984	0.980	0.940	0.72	0.998	0.994	0.986	0.964

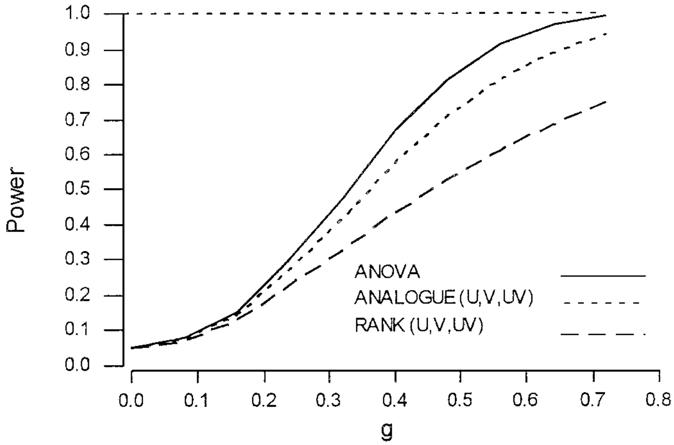

Fig. 2a. Comparison of powers of ANOVA, RANK and ANALOGUE for testing the linear component of U for the model (U, V, UV) with the design $(4, 4, 2)$

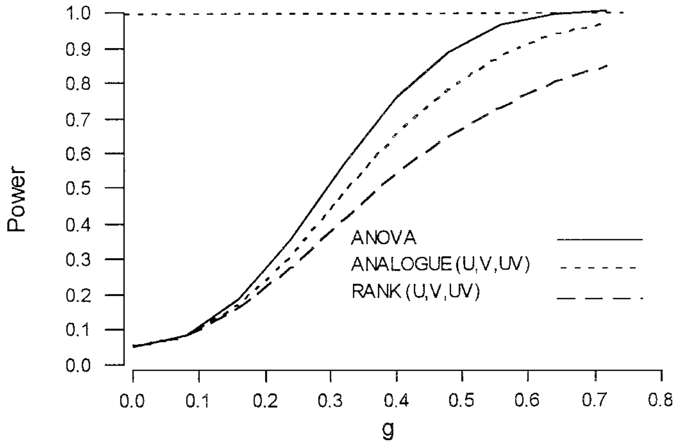

Fig. 2b. Comparison of powers of ANOVA, RANK and ANALOGUE for testing the linear by linear component of the interaction UV for the model (U, V, UV) with the design $(4, 4, 2)$

7 Discussion

This work was motivated by the requirements of a practical problem involving a blocked factorial experiment in which the responses were available as rankings within blocks. Initially a ranking method based on partitioning Friedman's test statistic into $K - 1$ 'orthogonal' components, representing the linear, quadratic, etc, polynomials for each factor and the linear by linear etc, components of the interaction, was proposed. This procedure is equivalent to carrying out an analysis of variance on the Friedman ranks and partitioning the treatment sum of squares into its $K - 1$ components with contrast squares

T_l^2/M for $l = 1, \ldots, K-1$, each with one degree of freedom, and then comparing these with the mean square based on the total sum of squares

$$\sum_{m=1}^{M}\sum_{k=1}^{K} R_{mk}^2 - \frac{\left(\sum_{m=1}^{M}\sum_{k=1}^{K} R_{mk}\right)^2}{MK} = \frac{MK(K^2-1)}{12},$$

which has $M(K-1)$ degrees of freedom because the ranks are constrained to sum to $K(K+1)/2$ within each block.

The problem encountered with this form of the test is that this mean square is inflated under any alternative hypothesis, so that there is a loss of power associated with testing one component when other effects are present in the model. To overcome this problem, an alternative procedure is introduced which involves comparing the contrast squares with the residual mean square from the same analysis of variance. It is shown that the power losses are very much reduced with this modification. There is also the bonus that the required analysis is simply obtained, since it is the same analysis of variance that would have been applied to the normal data if these had been available.

There are, however, two possible disadvantages with this method. The first is that use of the asymptotic distributions will result in increased probability of type I errors for very small designs as discussed in Section 5.1. However, this should not be a problem in most practical situations, where at least one of the factors has three or more levels or where the experiment involves more than two blocks. The second problem is that the residual mean square can be zero if the rankings are identical within each block. This can occur even in the null situation with very small designs and could occur in larger designs when the alternative model contains very pronounced differences over the levels of both factors. This latter situation did not arise in any of the simulations used in this study except when the designs were small.

Although the contrasts used in the tests are uncorrelated under the null hypothesis of no main effects and interactions, it was the case that the performance of the test for one component was affected by the presence of other components in the model. Since we apply ranks within blocks, we did not observe the disadvantageous feature of the rank transform test for interaction discussed by Thompson (1991). He warns that the test for interaction based on the rank transform applied to all the data, can have a large type I error rate, even for large samples, when certain main effects are present.

Further studies are underway, theoretically and through simulations, to investigate the properties of the proposed ANALOGUE ranking method under the null situation, with various alternative models and with different non-normal error structures.

References

Akritas MG (1990) The rank transform method in some two-factor designs. Journal of the American Statistical Association, 85:73–78

Barlow RE, Bartholomew DJ, Brenner JM, Brunk HD (1972) Statistical inference under order restrictions. Wiley, New York

Blair RC, Sawilowsky SS, Higgins JJ (1987) Limitations of the rank transform statistics in tests for interaction. Communications in Statistics – Simulation and Computation 16:1133–1145

Brunner E, Neumann N (1986) Rank tests in 2 × 2 designs. Statistica Neerlandica 40:251–272
Conover WJ, Iman RL (1976) On some alternative procedure using ranks for the analysis of experimental designs. Communications in Statistics – Theory and Methods A5:1349–1368
Hollander M (1967) Rank tests for randomised blocks when the alternatives have an a priori ordering. Annals of Mathematical Statistics 38:867–877
Iman RL (1974) A power study of a rank transform for the two-way classification model when interactions may be present. Canadian Journal of Statistics 2:227–239
Jonckheere AR (1954) A distribution-free k-sample test against ordered alternatives. Biometrika 41:133–145
Lemmer HH (1980) Some empirical results on the two-way analysis of variance by ranks. Communications in Statistics – Theory and Methods A9:1427–1438
Puri ML, Sen PK (1968) On Chernoff-Savage tests for ordered alternatives in randomised blocks. Annals of Mathematical Statistics 39:967–972
Sawilowsky SS, Blair RC, Higgins JJ (1989) An investigation of type I error and power properties of the rank transform procedure in factorial ANOVA. Journal of Educational Statistics 14:255–267
Scheirer CJ, Ray WS, Hare N (1976) The analysis of ranked data derived from completely randomised factorial designs. Biometrics 32:429–434
Shirley EAC (1987) Applications of ranking methods to multiple comparison procedures and factorial experiments. Applied Statistics 36:205–213
Shorack GR (1967) Testing against ordered alternatives in model I analysis of variance: normal theory and non-parametric. Annals of Mathematical Statistics 38:1740–1752
Skillings JH, Wolfe DA (1978) Distribution-free tests for ordinal alternatives in a randomized block design. Journal of the American Statistical Association 73:427–431
Steel RGD (1960) A rank sum test for comparing all pairs of treatments. Technometrics 2:197–207
Terpstra TJ (1953) The exact probability distribution of the t-statistic for testing against trend and its normal approximation. Proceedings Koninklijke Nederlandse Akademie van Wetenschappen 56:433–437
Thomas GE, Kiwanga SS (1993) Use of ranking and scoring methods in the analysis of ordered categorical data from factorial experiments. The Statistician 42:55–67
Thompson GL (1991) A note on the rank transform for interactions. Biometrika 78:697–701
Tryon PV, Hettmansperger TP (1974) A class of non-parametric tests for homogeneity against ordered alternatives. Annals of Statistics 1:1061–1070
Williams DA (1986) A note on Shirley's non-parametric test for comparing several dose levels with a zero-dose control. Biometrics 42:183–186