Using Weighted Rankings in the Analysis of Complete Blocks With Additive Block Effects

DANA QUADE*

The standard nonparametric procedures for testing the hypothesis of no treatment effects in a complete blocks experiment depend entirely on the within-block rankings. If block effects are assumed additive, however, then between-block information may be recovered by weighting these rankings according to their credibility with respect to treatment ordering. (For the special case of only two treatments, the sign test exemplifies use of unweighted rankings and the signed-rank test weighted.) A general family of weighted-rankings test statistics for comparing two or more treatments is presented. They are simple to compute, are strictly distribution free, and have asymptotic chi-squared distributions.

KEY WORDS: Ranks; Weighted rankings; Complete blocks.

1. INTRODUCTION

Let X_{ij} be the observation on the jth of m treatments in the ith of n complete blocks, and consider testing the hypothesis of no treatment effects, specifically

$$H_0\colon X_{i1}, \ldots, X_{im} \text{ are interchangeable for each } i\ .$$

(By definition, random variables are interchangeable if their joint distribution function is invariant under permutations.) The alternatives under consideration are fairly general; however, a specific example is that of additive treatment effects:

H_1: There exist quantities $\tau_1, \ldots, \tau_m$ (treatment effects), not all equal to zero, such that for $i = 1, \ldots, n$, $X_{i1} - \tau_1, \ldots, X_{im} - \tau_m$ are interchangeable.

Standard nonparametric procedures for attacking this problem, for example, the tests of Friedman (1937) and of Brown and Mood (1951), are based on within-block rankings. Thus, the only assumption they require is that the blocks be independent.

Assumption 1: The random vectors

$$\mathbf{X}_i = (X_{i1}, \ldots, X_{im})'\ ,$$

for $i = 1, \ldots, n$ (i.e., the blocks), are mutually independent.

It is common to make the additional assumption of additive block effects.

Assumption 2: There exist quantities $\beta_1, \ldots, \beta_n$ (block effects) such that the random vectors $(X_{i1} - \beta_i, \ldots, X_{im} - \beta_i)'$ are all identically distributed.

With this assumption, comparisons of observations are possible between blocks as well as within. Thus, procedures that use only within-block comparisons waste information. For further discussion of this point, see Hodges and Lehmann (1962).

With $m = 2$ treatments only, the use of within-block rankings is equivalent to performing a sign test. For this special case, however, a simple distribution-free procedure is available that recovers between-block information: the signed-rank test of Wilcoxon (1945). For $m \geq 3$, the situation is less satisfactory. A permutation test, based on the classical two-way analysis of variance statistic, was proposed by Pitman (1938), but the calculations it requires are so extensive as to render it infeasible except for small m and n. Another procedure, originally proposed by Hodges and Lehmann (1962), involves analyzing the observations after first aligning them by subtracting out estimates of the block effects, but this is only asymptotically distribution free, except insofar as it also is a permutation test. Finally, a procedure of Doksum (1967) consists essentially of considering signed-rank tests on all pairs of treatments simultaneously, but this too is only asymptotically distribution free. Thus, there is still need for a method that recovers between-block information for $m \geq 3$ while remaining both feasible to compute and strictly distribution free.

2. WEIGHTED RANKINGS

In this section is presented a class of procedures based on a method of weighted within-block rankings that generalizes the standard method based on unweighted rankings. The intuitive idea behind these procedures seems to have been first expressed in a rarely cited article by Tukey (1957). Suppose the observations on different treatments are more distinct in some blocks than in the others; then it seems intuitively reasonable that the ordering of the treatments that these blocks suggest is more likely to reflect the underlying true ordering. These same blocks might more or less equivalently be described as having greater observed variability, although the word *observed* is to be emphasized because by Assumption 2 all blocks are identically distributed except for additive block effects. Thus, these blocks, which may be referred

* Dana Quade is Professor, Department of Biostatistics, School of Public Health, University of North Carolina, Chapel Hill, NC 27514. This work was partly supported by a Research Career Development Award (No. GM-38906) from the National Institute of General Medical Sciences.

September 1979, Volume 74, Number 367
Theory and Methods Section

to as more credible with respect to treatment ordering, will be given greater weight in the analysis.

Consider in more detail the structure of the test statistics based on the standard method. For simplicity of exposition, make the (unessential) assumption

Assumption 3:

$$P\{X_{ij} = X_{ij'}\} = 0 \quad \text{for} \quad j \neq j' ,$$

so that there will be no within-block ties, and let R_{ij} be the within-block rank of X_{ij}: $(R_{ij}, \ldots, R_{im})$ is then a permutation of $(1, \ldots, m)$. Let $t_1, \ldots, t_m$ be a fixed set of treatment scores, not all equal. Define $\bar{t} = \sum t_j/m$ and $T = \sum (t_j - \bar{t})^2$. Then,

$$C_{ii'} = \sum_{j=1}^{m} (t_{R_{ij}} - \bar{t})(t_{R_{i'j}} - \bar{t})/T$$

is a measure of rank correlation between the ith and i'th blocks; in particular, if $t_j = j$ this is the Spearman rank correlation. A measure of agreement among the blocks is the average rank correlation

$$C^* = \sum\sum_{i \neq i'} C_{ii'}/n(n-1) ,$$

and the corresponding test statistic is

$$W = (m-1)[1 + (n-1)C^*] = (m-1) \sum_{j=1}^{m} [\sum_{i=1}^{n} (t_{R_{ij}} - \bar{t})]^2/nT .$$

Sen (1968) shows that as n tends to infinity, W has asymptotically a χ^2 distribution with $(m-1)$ degrees of freedom.

To determine the weight for the ith block, use some location-free statistic $D_i = D(X_{i1}, \ldots, X_{im})$ that measures the credibility of the block with respect to treatment ordering, and let Q_i be the rank of D_i among $D_1, \ldots, D_n$. Again, for simplicity of exposition, make the (unessential) assumption

Assumption 4:

$$P\{D_i = D_{i'}\} = 0 \quad \text{for} \quad i \neq i' ,$$

so that there will be no ties in the ranking of the blocks. Now, let $b_1, \ldots, b_n$ be a fixed set of block scores, and define $B_k = \sum b_i^k$. Then the weight given to the ith block will be proportional to b_{Q_i}, and the weight given to the correlation between the ith and i'th blocks proportional to $b_{Q_i}b_{Q_{i'}}$, so that the weighted average rank correlation becomes

$$C = \sum\sum_{i \neq i'} b_{Q_i}b_{Q_{i'}}C_{ii'}/(B_1^2 - B_2) ,$$

and the test statistic is

$$W = (m-1)[1 + (B_1^2/B_2 - 1)C] = (m-1) \sum_{j=1}^{m} [\sum_{i=1}^{n} b_{Q_i}(t_{R_{ij}} - \bar{t})]^2/B_2T .$$

Given any particular choice of the block scores b_i and the treatment scores t_j, there must be some block in which the observations are replaced by $b_1t_1, \ldots, b_1t_m$ in some order, another block with $b_2t_1, \ldots, b_2t_m$, and so on to $b_nt_1, \ldots, b_nt_m$. But all rankings of the blocks are equally likely because the statistics D_i are independent (by Assumption 1) and identically distributed (by Assumption 3), and all rankings within the blocks are equally likely under the null hypothesis. Hence the weighted-rankings statistic is strictly distribution free under H_0 (and independent of the choice of the credibility measure D); thus it could be tabulated, at least for small m and n.

Now consider the asymptotic distribution of W as the number of blocks increases without limit.

Theorem: Suppose Assumptions 1 and 2 are satisfied (make Assumptions 3 and 4 also, to simplify the exposition by preventing ties, but they are not really essential). Then under H_0 as the number of blocks tends to infinity the test statistic W has asymptotically a $\chi^2(m-1)$ distribution, provided that

$$\sum_i (b_{ni} - \bar{b}_n)^r/[\sum (b_{ni} - \bar{b}_n)^2]^{r/2} = O(n^{1-r/2}) \quad \text{for} \quad r = 3, 4, \ldots ,$$

where $\bar{b}_n = \sum b_{ni}/n$ (and n has been indicated explicitly by writing b_{ni} for b_i—elsewhere n is usually suppressed).

Proof: The theorems referred to as (3.4.1), (3.4.5), and (7.2.1) are from Puri and Sen (1971). Define treatment totals $G_j = \sum b_{nQ_i}t_{R_{ij}}$ for $j = 1, \ldots, m$, and consider an arbitrary contrast in them, say $G = \sum g_jG_j$ where $\sum g_j = 0$ and $\sum g_j^2 > 0$. Then, $G = \sum b_{nQ_i}A_i$ where $A_i = \sum g_jt_{R_{ij}}$. By Assumptions 1 and 2 the A_i's are independent and identically distributed random variables, and they are clearly bounded. Under H_0 they have mean 0 and (using Assumption 3) variance $T\sum g_j^2/(m-1) > 0$, so they satisfy the conditions of (3.4.5). Furthermore, by Assumptions 1 and 2 (and 4), the Q_i's are a random permutation of the integers $1, \ldots, n$; under H_0 they are independent of the A_i's. Hence, it is easy to verify that G has mean 0 and variance

$$\sigma^2 = (\sum b_{ni}^2)(\sum t_j^2)(\sum g_j^2)/(m-1) .$$

Thus, by (3.4.1) the ratio G/σ is asymptotically a standard normal variable, and the present theorem follows using the same argument as for (7.2.1).

Simple but tedious algebra yields the lower moments of W under H_0:

$$E\{W\} = m - 1 ,$$
$$E[W - (m-1)]^2 = 2(m-1)\gamma_{1n} ,$$
$$E[W - (m-1)]^3 = 8(m-1)\{\gamma_{2n} + \gamma_{3n}\gamma_m\} ,$$

where

$$\gamma_{1n} = 1 - B_4/B_2^2 ,$$
$$\gamma_{2n} = 1 - 3B_4/B_2^2 + 2B_6/B_2^3$$
$$\gamma_{3n} = (B_3^2 - B_6)/B_2^3 ,$$

and

$$\gamma_m = m(m-1)[\sum (t_i - \bar{t})^3]^2/2(m-2)T^3 .$$

Note that $\gamma_m = 0$ if the t's are symmetrically placed about their median, and that, under the condition of the theorem, γ_{1n} and γ_{2n} tend to 1 and γ_{3n} to 0 as $n \to \infty$. It then follows that

$$[W - (m-1)]\gamma_{1n}/(\gamma_{2n} + \gamma_{3n}\gamma_m) + \delta$$

may be approximated by a χ^2 with

$$\delta = (m-1)\gamma_{1n}^3/(\gamma_{2n} + \gamma_{3n}\gamma_m)^2$$

degrees of freedom. This approximation improves on the one given by the theorem in that it fits three moments exactly, rather than only one, for all values of n.

3. WHICH WEIGHTED-RANKINGS STATISTIC?

There are several dimensions of choice available in the class of weighted-rankings statistics as defined earlier: We must decide on the treatment scores t_j, the block scores b_i, and the credibility measure D. Recalling that under H_0 statistics of the form of W all have expectation $(m-1)$, a reasonable approach is to ask which choices would maximize the expectation under H_1. Define

$$\theta_{kj} = E[t_{R_{ij}} - \bar{t} \mid Q_i = k] \ ;$$

this may be described in words as the expected value of the score for the jth treatment, corrected for the mean, in the kth-most credible block. Then,

$$E[C_{ii'} \mid Q_i = k, Q_{i'} = k'] = \sum_{j=1}^{m} \theta_{kj}\theta_{k'j}/T \ ,$$

and

$$E[C] = E[2b_{Q_i}b_{Q_{i'}}C_{ii'}/(B_1^2 - B_2)]$$
$$= \sum_{j=1}^{m} \sum\sum_{k \neq k'} b_k b_{k'} \theta_{kj}\theta_{k'j}/(B_1^2 - B_2)T \ .$$

Hence,

$$E[W] = (m-1)[1 + \sum_{j=1}^{m} \{(\sum_{k=1}^{n} b_k\theta_{kj})^2 - \sum_{k=1}^{m} b_k^2\theta_{kj}^2\}/B_2T] \ .$$

or, in matrix notation,

$$E[W] = (m-1)[1 + \mathbf{b}'(\Theta\Theta' - \Delta)\mathbf{b}/\mathbf{b}'\mathbf{b}T] \ ,$$

where $\mathbf{b} = (b_1, \ldots, b_n)'$, $\Theta = ((\theta_{ij}))$, and

$$\Delta = \operatorname{diag}(\sum_j \theta_{ij}^2) \ .$$

Thus, if $\mathbf{b}$ is chosen as the characteristic vector corresponding to the largest characteristic root (say, λ_1) of $(\Theta\Theta' - \Delta)$, then $E[W] = (m-1)[1 + \lambda_1/T]$, and this is its maximum possible value.

Define

$$\psi_j = \sum_k \theta_{kj}/n = E[t_{R_{ij}} - \bar{t}]$$

(note $\sum \psi_j = 0$); and, assuming $\Psi = \sum \psi_j^2 > 0$, define

$$\phi_k = \sum_j \psi_j\theta_{kj}/\Psi$$

(note $\sum \phi_k = n$). If the credibility measure D has been well chosen, then $0 \le \phi_1 \le \ldots \le \phi_n$; if it is irrelevant, then $\phi_k = 1$ for all k. Now consider approximating θ_{kj} by $\theta_k\psi_j$. With this approximation, $\Theta\Theta' = \Psi\boldsymbol{\phi}\boldsymbol{\phi}'$ where $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_n)'$, and $\Delta = \Psi \operatorname{diag}(\phi_i^2)$, so

$$E[W] = (m-1)[1 + \Psi\mathbf{b}'(\boldsymbol{\phi}\boldsymbol{\phi}' - \operatorname{diag}(\phi_i^2))\mathbf{b}/\mathbf{b}'\mathbf{b}T] \ .$$

For large n the matrix $\operatorname{diag}(\phi_i^2)$ can be neglected, and the optimal $\mathbf{b} = \boldsymbol{\phi}$, the characteristic vector corresponding to the largest characteristic root $\boldsymbol{\phi}'\boldsymbol{\phi}$ of $\boldsymbol{\phi}\boldsymbol{\phi}'$; then

$$E[W] = (m-1)[1 + \Psi\boldsymbol{\phi}'\boldsymbol{\phi}/T] \ .$$

With unweighted rankings, $\mathbf{b} = (1, \ldots, 1)'$, and

$$E[W] = (m-1)[1 + n\Psi/T] \ .$$

Note that $\boldsymbol{\phi}'\boldsymbol{\phi} \ge n$, suggesting greater efficiency for weighted rankings.

The approach just outlined suggests that the optimal treatment scores are the same for the weighted-rankings situation as for the unweighted situation discussed by Sen (1968). In what follows, only simple linear scores $t_j = j$ are used. Unfortunately, no method is yet available for calculating the expectations θ_{kj}, and hence no analytic solution to the problem of choosing block scores b_i or credibility measure D can be provided. Instead, one must rely on intuitive notions, with some guidance from Monte Carlo work.

A particularly simple choice is to take zero-one block scores

$$b_i = 0 \quad \text{if} \quad i = 1, \ldots, l$$
$$= 1 \quad \text{if} \quad i = l+1, \ldots, n \ .$$

This amounts to discarding from analysis the l lowest-ranking blocks. The great advantage of this choice is that it allows use of existing tables. With these scores, the null-hypothesis distribution of the weighted statistic for n blocks is the same as that of the corresponding unweighted statistic for $(n - l)$ blocks.

Another simple and intuitive choice—the one suggested by Tukey (1957)—is to use linear scores

$$b_i = i \ , \quad i = 1, \ldots, n \ .$$

These scores directly generalize the signed-rank statistic to $m \ge 3$. With linear scores for both blocks and treatments, the test statistic becomes

$$W = \frac{72S}{m(m+1)n(n+1)(2n+1)} - \frac{9(m+1)n(n+1)}{2(2n+1)} \ ,$$

where

$$S = \sum_{j=1}^{m} [\sum_{i=1}^{n} Q_i R_{ij}]^2$$

is a convenient integer. Quade (1972) provides tables of the exact null-hypothesis distribution of W in this special case for the following combinations of m and n: (3, 3), (3, 4), (3, 5), (3, 6), (3, 7), (4, 3), (4, 4), and (5, 3). Because the successive values of S differ by 2 (or multiples of 2), a continuity correction of 1 to S may be employed in applying the asymptotic χ^2 approximations. Also, the quantities B_k are given by well-known formulas for sums of powers of integers, where in particular one obtains

$$\gamma_{1n} = 1 - 6(3n^2 + 3n - 1)/5n(n + 1)(2n + 1) \ ,$$
$$\gamma_{2n} = 3\gamma_{1n} - 2 + 72(3n^4 + 6n^3 - 3n + 1)/7n^2(n + 1)^2(2n + 1)^2 \ .$$

Thus, $\{[W - (m - 1)]\gamma_{1n}/\gamma_{2n} + \delta\}$ is distributed as $\chi^2(\delta)$ where $\delta = (m - 1)\gamma_{1n}^3/\gamma_{2n}^2$.

For his doctoral dissertation, Silva (1977) made an extensive Monte Carlo investigation into the efficiency of weighted-rankings procedures. His results indicate considerable advantages for linearly weighted rankings over unweighted, both in small experiments and asymptotically for large n, with normal or uniform errors. Results were mixed with double-exponential errors. Zero-one block scores also were considered, but performed poorly. (Linear treatment scores were used exclusively.)

Finally, with respect to the credibility measure D, Silva's Monte Carlo work suggests that the choice may not be crucial. He considered the range, standard deviation, mean deviation, interquartile difference, and mean difference, but found at most trivial differences among them. Thus, the range may be tentatively recommended on the grounds of simplicity. Because the work by Tukey (1957) did not come to attention until later, his suggestion to use the least difference between any two observations within a block was not evaluated.

The following little example, chosen for ease of calculation, may help readers to check their understanding of the formulas. Suppose there are $m = 3$ treatments and $n = 7$ blocks, with raw data (X_{ij}) as follows:

Block		1	2	3	4	5	6	7
	A	52	63	45	53	47	62	49
Treatment	B	45	79	57	51	50	72	52
	C	38	50	39	43	56	49	40

Standard procedures include ordinary two-way analysis of variance, which yields for these data $F(2, 12) = 6.841$ and hence $P = .010$, and Friedman's test, for which $\chi^2(2) = 6.000$ and the exact $P = .051$. The block ranking, using ranges or standard deviations, is 4752163. Discarding the two lowest-ranking blocks, and thereby reducing n from 7 to 5, produces Friedman's $\chi^2(2) = 8.400$ and exact $P = .008$. Using linear block scores yields $S = 10{,}550$ and $W = 8.157$ and, thus, exact $P = .005$ according to Quade (1972). With continuity correction, $W = 8.150$, and taking this as $\chi^2(2)$ gives approximate $P = .017$. For the more complicated three-moment approximation, $\gamma_1 = .761$ and $\gamma_2 = .419$, and then $\chi^2(5.029) = 16.205$, giving the more accurate approximation $P = .006$.

4. CONCLUDING REMARKS

Let the reader be reminded that all blocks are assumed to have equal underlying variability, so that those with greater observed variability are more credible with respect to treatment ordering. If instead it is suspected that blocks with greater observed variability may have greater underlying variability, then one should perhaps weight them less rather than more: Indeed, it may then be grossly inefficient to use the method of weighted rankings (or any of its competitors, including ordinary analysis of variance). See Skillings (1978) for work along these lines.

Finally, in this article attention has been restricted to experiments with exactly one observation per treatment per block. Silva (1977) has extended the method to balanced incomplete blocks, and it is conceptually simple to extend to other balanced cases. Silva also has considered extension to groups of experiments with additive block effects within groups but more general block differences between groups.

[*Received January 1978. Revised March 1979.*]

REFERENCES

Brown, G.W., and Mood, A.M. (1951), "On Median Tests for Linear Hypotheses," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley: University of California Press, 159–166.

Doksum, K.A. (1967), "Robust Procedures for Some Linear Models With One Observation per Cell," *Annals of Mathematical Statistics*, 38, 878–883.

Friedman, Milton (1937), "The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance," *Journal of the American Statistical Association*, 32, 675–701.

Hodges, J.L., Jr., and Lehmann, E.L. (1962), "Rank Methods for Combination of Independent Experiments in Analysis of Variance," *Annals of Mathematical Statistics*, 33, 482–497.

Pitman, E.J.G. (1938), "Significance Tests Which May Be Applied to Samples From Any Population, III. The Analysis of Variance Test," *Biometrika*, 29, 322–335.

Puri, M.L., and Sen, P.K. (1971), *Nonparametric Methods in Multivariate Analysis*, New York: John Wiley & Sons.

Quade, Dana (1972), "Analyzing Randomized Blocks by Weighted Rankings," Report SW 18/72 of the Mathematical Center, Amsterdam.

Sen, P.K. (1968), "Asymptotically Efficient Tests by the Method of n Rankings," *Journal of the Royal Statistical Society*, Ser. B, 30, 312–317.

Silva, Claudio (1977), "Analysis of Randomized Blocks Designs Based on Weighted Rankings," Chapel Hill, N.C.: Institute of Statistics Mimeo Series No. 1137.

Skillings, J.H. (1978), "Adaptively Combining Independent Jonckheere Statistics in a Randomized Block Design With Unequal Scales," *Communications in Statistics*, A7, 1027–1039.

Tukey, J.W. (1957), "Sums of Random Partitions of Ranks," *Annals of Mathematical Statistics*, 28, 987–992.

Wilcoxon, Frank (1945), "Individual Comparisons by Ranking Methods," *Biometrics*, 1, 80–83.