

AD-A135-759

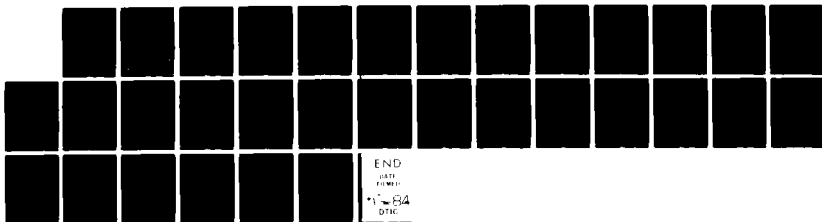
TESTING HYPOTHESES IN LINEAR MODELS WITH WEIGHTED RANK
STATISTICS(U) WESTERN MICHIGAN UNIV KALAMAZOO DEPT OF
MATHEMATICS G L SIEVERS NOV 83 TR-71 N00014-78-C-0637

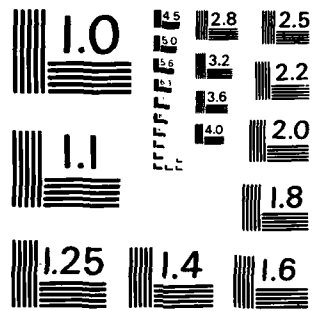
1/1

UNCLASSIFIED

F/G 12/1

NL

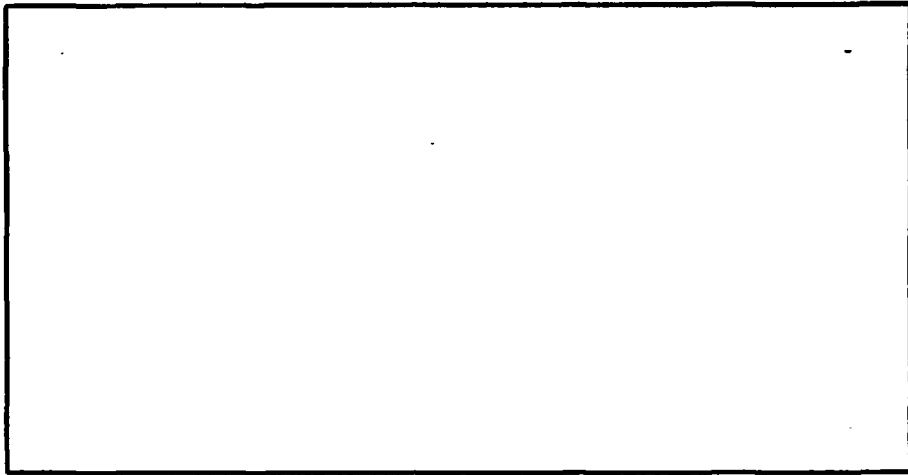




MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(16)

AD-A135 759



DEPARTMENT OF MATHEMATICS

WESTERN MICHIGAN UNIVERSITY

Kalamazoo, Michigan

49008

DTIC
SELECTED
DEC 14 1983
A

DTIC FILE COPY

This document has been approved
for public release and sale; its
distribution is unlimited.

83 12 13 300

1

TESTING HYPOTHESES IN LINEAR MODELS
WITH WEIGHTED RANK STATISTICS

by

Gerald L. Sievers

TECHNICAL REPORT NO. 71
November, 1983

This research was partially supported by the Office of Naval Research
under contract N00014-78-C-0637 (NR 042-407)

Reproduction in whole or in part is permitted
for any purpose of the United States Government

Approved for public release; distribution unlimited

DEPARTMENT OF MATHEMATICS
WESTERN MICHIGAN UNIVERSITY
KALAMAZOO, MICHIGAN 49008

ABSTRACT

Tests of hypotheses for the parameters in a general linear model are considered based on weighted rank statistics. Results are presented for tests based on a rank estimate, tests based on drop in dispersion and aligned rank tests. Weights can be used to focus the analysis on simple effects and provide an additional degree of robustness to rank tests. Several analysis of variance applications are discussed.

Key Words and Phrases: analysis of variance, dispersion function, robust tests, aligned ranks, rank estimates, weights



A-1

1. INTRODUCTION

Statistical procedures based on ranks are widely used for simple linear model problems. For the general linear model there has also been considerable development in the area of rank methods. The results show that they are nearly as efficient as the classical, least-squares methods when normal distributions hold and more efficient for many other distributions. The least-squares methods can be inefficient for non-normal distributions (for both large and small sample sizes) and are sensitive to outliers and high leverage points, while the rank methods are more robust.

This paper will consider weighted rank statistics for linear model problems. Weights are usually introduced in statistical methods to increase efficiency but that is not the case here. Instead, the interest is in using weights so as to not lose efficiency while gaining in other respects, in particular, in gaining a further degree of robustness. This is discussed further in section 6.

The emphasis will be on analysis of variance problems and in this case weights of zero or one are of special concern. They have the effect of restricting the ranking to various subsets of the data instead of ranking the entire set. A familiar example is the within block ranking used for Friedman's test. It should be noted that the large sample results considered here would apply in block designs with within block comparisons as the number of observations per block grows large with a fixed number of blocks and they would not apply in the reverse case of fixed block size with the number of blocks growing large.

In one-way and higher order analysis of variance problems there are many rank tests in popular use that are based on restricted ranking, restricted comparison methods. For example, in block designs the ranking may be done separately in each block with no comparison between observations in different blocks. In testing the effects of several treatments against a control, the treatment groups may be compared only to the control group and not to each other. In testing the equality of several groups against an ordered alternative Tryon and Hettmansperger (1973) discussed the value of using only comparisons between adjacent groups. The methodology of this paper includes the types of restricted comparisons above as special cases.

Results on the asymptotic distributions of estimates and test statistics that have appeared separately in the literature can be obtained in a unified framework from Remarks 2.1 and 2.3. This common view of many diverse problems can be valuable in promoting understanding of their common structure and in suggesting new solutions to other problems. The general results may provide the details necessary to modify a standard procedure to better fit a particular application. A computer program based on the general approach would be able to handle a wide range of problems.

Consider the linear model

$$(1.1) \quad \underline{Y} = \beta_0 \underline{1} + \underline{X}\beta + \underline{e},$$

where $\underline{Y} = (Y_1, \dots, Y_n)'$ is an $n \times 1$ random vector, $\underline{1}$ is an $n \times 1$ vector with each element equal to one, $\underline{X} = (x_{ij})$ is an $n \times p$ design matrix, $\beta = (\beta_1, \dots, \beta_p)'$ is a $p \times 1$ vector of parameters and

$\underline{e} = (e_1, \dots, e_n)'$ is an $n \times 1$ vector of random errors. Assume that \underline{X} is centered so that its column sums are zero. Assume that the errors are independent with a common distribution having density function f . The residuals are denoted by $\underline{z} = (z_1, \dots, z_n)'$ where

$$\underline{z} = \underline{z}(\underline{b}) = \underline{Y} - \underline{X}\underline{b}.$$

A weighted rank estimate of $\underline{\beta}$ is reviewed in section 2 along with some theoretical results that will be needed in the rest of the paper. Longer proofs are delayed until the Appendix. Sections 3, 4 and 5 discuss tests of hypotheses based on the estimate $\hat{\underline{\beta}}$, on the drop in dispersion and on aligned ranks, respectively. Some advantages in the use of weights are considered in section 6 and applications to a variety of analysis of variance problems are discussed in section 7.

2. PRELIMINARIES

Consider the dispersion function of the residuals

$$(2.1) \quad D = D(b) = \sum_{i < j} w_{ij} |Z_i - Z_j|,$$

where the $w_{ij} \geq 0$, $1 \leq i < j \leq n$ are a given set of weights. The weights should reflect the importance of the comparisons. They may depend on the design matrix X . Some of the weights can be zero to drop some pairwise comparisons from consideration. The special case of equal weights, $w_{ij} \equiv 1$, gives rise to Gini's mean difference. Hettmansperger and McKean (1978a) have shown that in this case the dispersion function is equivalent to Jaeckel's dispersion function with Wilcoxon scores.

The dispersion function D can be expressed in another form. Let (R_1, \dots, R_n) denote the ranks of the residuals; that is, R_i is the rank of Z_i in the set $\{Z_1, \dots, Z_n\}$, $1 \leq i \leq n$. Let $\text{sgn}(v) = +1, 0, -1$ as v is $> 0, = 0, < 0$. Extend the definition of the weights w_{ij} to all subscripts $i, j = 1, \dots, n$ by using $w_{ji} = w_{ij}$ and $w_{ii} = 0$. Then, using $|v| = v \text{sgn}(v)$, some manipulation shows

$$(2.2) \quad D = \sum_{i=1}^n B_i Z_i,$$

with $B_i = B_i(b) = \sum_j w_{ij} \text{sgn}(Z_i - Z_j)$, $i = 1, \dots, n$. The coefficients B_i are random with B_i depending on the rank of Z_i and also on the subscripts of the residuals that are less than Z_i . In the special case $w_{ij} \equiv 1$, $B_i = 2R_i - (n+1)$.

The estimate $\hat{\beta}$ of the parameter β will be a point in the parameter space minimizing the dispersion function (see Sievers (1983)). The partial derivatives of D are (approximately) equal to zero at the minimum. Using (2.2), these derivatives are

$$(2.3) \quad \partial D / \partial b_k = - \sum_{i=1}^n B_i x_{ik}$$

for $k = 1, \dots, p$, except at a finite number of points. Letting $a_{ij}(k) = w_{ij}(x_{jk} - x_{ik})$, another form of the derivatives is

$$(2.4) \quad \partial D / \partial b_k = -2 \sum_{i < j} a_{ij}(k) \phi(Z_i, Z_j) + \sum_{i < j} a_{ij}(k),$$

where $\phi(u, v) = 0, \frac{1}{2}, 1$ as $u > v, u = v, u < v$. A matrix form is

$$(2.5) \quad \partial D / \partial \underline{b} = -\underline{X}' \underline{B},$$

where $\underline{B} = (B_i)$ is $n \times 1$.

For the results to follow it is convenient to use a multiple of the derivative. Define a random vector $\underline{U}(\underline{b}) = (U_1(\underline{b}), \dots, U_p(\underline{b}))'$ by

$$(2.6) \quad \underline{U}(\underline{b}) = (1/2)n^{-3/2} \underline{X}' \underline{B}.$$

Note that $U_k(\underline{b}) = n^{-3/2} [\sum_{i < j} a_{ij}(k) \phi(Z_i, Z_j) - \sum_{i < j} a_{ij}(k)/2]$.

Some constants will also be needed. For $k = 1, \dots, p$, let

$$a_{i.}(k) = \sum_{j=i+1}^n a_{ij}(k) \quad \text{for } i = 1, \dots, n-1,$$

$$a_{.j}(k) = \sum_{i=1}^{j-1} a_{ij}(k) \quad \text{for } j = 2, \dots, n,$$

$$a_{n.}(k) = 0, \quad a_{.1}(k) = 0, \quad a_{..}(k) = \sum_{i < j} a_{ij}(k) \quad \text{and}$$

$$A_i(k) = a_{.i}(k) - a_{i.}(k).$$

For asymptotic purposes, a sequence of these constants is needed, indexed on $n = 1, 2, \dots$, but this dependence on n for these and other quantities will sometimes be suppressed in the notation.

Let \underline{W}_n be an $n \times n$ symmetric matrix involving the weights of the dispersion function. Specifically, define the $(i, j)^{\text{th}}$ element of \underline{W}_n to be $-w_{ij}$ if $i < j$ and $-w_{ji}$ if $i > j$. The i^{th} diagonal element of \underline{W}_n is $w_i = \sum_{j \neq i} w_{ij}$. Thus \underline{W}_n has the negatives of the dispersion function weights for its off-diagonal elements and positive diagonal elements determined so that the row and column sums are zero.

Also write

$$\underline{A}_n = \underline{W}_n \underline{X},$$

$$\underline{V}_n = \underline{X}' \underline{W}_n \underline{W}_n \underline{X}.$$

$$\underline{C}_n = \underline{X}' \underline{W}_n \underline{X}.$$

ASSUMPTION (A₁): For each $k = 1, \dots, p$

$$\sum_{i=1}^n A_i^2(k) / \max_{1 \leq i \leq n} A_i^2(k) \rightarrow \infty \quad \text{as } n \rightarrow \infty.$$

ASSUMPTION (A₂): For each $k = 1, \dots, p$

$$\sum_{i < j} a_{ij}^2(k) / \sum_{i=1}^n A_i^2(k) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

ASSUMPTION (A₃): For each $k = 1, \dots, p$

$$\sum_{i < j} a_{ij}^2(k) / \binom{n}{2} \text{ is bounded as } n \rightarrow \infty.$$

ASSUMPTION (A₄): For each $k = 1, \dots, p$

$$\max_{1 \leq i \leq n} |x_{ik}| / \sqrt{n} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

ASSUMPTION (A₅):

$$(1/n) \underline{X}' \underline{X} \rightarrow \underline{\Sigma} \quad \text{as } n \rightarrow \infty,$$

where $\underline{\Sigma}$ is a $p \times p$, positive definite matrix.

ASSUMPTION (A₆):

$$n^{-3} \underline{V}_{-n} \rightarrow \underline{V} \quad \text{as } n \rightarrow \infty,$$

where \underline{V} is a $p \times p$, positive definite matrix.

ASSUMPTION (A₇):

$$n^{-2} \underline{C}_{-n} \rightarrow \underline{C} \quad \text{as } n \rightarrow \infty,$$

where \underline{C} is a $p \times p$, nonsingular matrix.

Let $G(y) = P(e_1 - e_2 \leq y)$ denote the cdf of the difference of independent random variables, each with density f .

ASSUMPTION (A₈): The cdf G has a density $g = G'$ and $g(y)$ is continuous at $y = 0$ with $g(0) > 0$.

ASSUMPTION (A₉): Assume the error density f is absolutely continuous and $\int (f'/f)^2 f \, dx < \infty$.

The following remarks are from Theorems 4.1, 5.1 and 6.1 of Sievers (1983) with some change in notation.

REMARK 2.1. Let $(A_1) - (A_9)$ hold. Let $\underline{\Delta}$ be a fixed $p \times 1$ vector. Then as $n \rightarrow \infty$,

$$\underline{U}(0) \Big|_{\underline{\Delta}/\sqrt{n}} \xrightarrow{L} N(\gamma \underline{C} \underline{\Delta}, (1/12) \underline{V}),$$

where $\gamma = (f f^2)$.

In the notation " $\Big|_{\underline{\beta}}$ ", the $\underline{\beta}$ specifies the parameter vector in model (1.1).

Let $c > 0$ be given and define a set $\mathcal{D} = \{\underline{\Delta} : -c \leq \Delta_k \leq c, k=1, \dots, p\}$. Let $\|\cdot\|$ be the Euclidean norm.

REMARK 2.2. Let $(A_1) - (A_9)$ hold. Then as $n \rightarrow \infty$,

$$\sup_{\underline{\Delta} \in \mathcal{D}} \left\| \underline{U}(\underline{\Delta}/\sqrt{n}) - \underline{U}(0) + \gamma \underline{C} \underline{\Delta} \right\| \Big|_0 \xrightarrow{P} 0.$$

REMARK 2.3. Let $(A_1) - (A_9)$ hold. Then as $n \rightarrow \infty$,

$$\sqrt{n}(\hat{\underline{\beta}} - \underline{\beta}) \Big|_{\underline{\beta}} \xrightarrow{L} N(0, (1/12\gamma^2) \underline{C}^{-1} \underline{V} \underline{C}^{-1}).$$

A test of the hypothesis concerning the full parameter vector, $H_0 : \underline{\beta} = \underline{0}$, can be based on a quadratic form in $\underline{U}(0)$ or in $\hat{\underline{\beta}}$ by using the large sample results in the preceding remarks. The former has the advantage of not requiring an estimate of γ .

3. TESTING A REDUCED MODEL

In this section the problem of testing a reduced model is discussed. With the full rank assumption on the design matrix, the problem will be expressed in terms of testing to drop some of the terms from the model. A test based on the rank estimate will be discussed here.

Consider the partitioning $\underline{X} = (\underline{X}_1, \underline{X}_2)$ and $\underline{\beta} = (\beta_1', \beta_2')$ where \underline{X}_1 is $n \times p_1$, \underline{X}_2 is $n \times p_2$, β_1 is $p_1 \times 1$, β_2 is $p_2 \times 1$ and $p_1 + p_2 = p$. The model (1.1) can then be written as

$$(3.1) \quad \underline{y} = \beta_0 \underline{1} + \underline{X}_1 \beta_1 + \underline{X}_2 \beta_2 + \underline{e}.$$

The reduced model hypothesis to be considered is $H_0: \beta_2 = \underline{0}$.

Some further notation will be needed. Let

$$\underline{C} = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}, \quad \underline{V} = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix},$$

where C_{11} and V_{11} are $p_1 \times p_1$, and let

$$\underline{G} = (-C_{21}C_{11}^{-1}, I_{p_2}), \quad C_{22 \cdot 1} = C_{22} - C_{21}C_{11}^{-1}C_{12}.$$

A natural test of the hypothesis H_0 can be based on a quadratic form in the estimate of β_2 . The estimate minimizing (2.1) can be partitioned $\hat{\underline{\beta}} = (\hat{\beta}_1', \hat{\beta}_2')$ and Remark 2.3 implies that

$$\sqrt{n}(\hat{\beta}_2 - \beta_2) |_{\beta} \xrightarrow{L} N(0, (1/12\gamma^2)\underline{\Sigma}_*),$$

where $\underline{\Sigma}_*$ is the lower right $p_2 \times p_2$ submatrix of $\underline{C}^{-1}\underline{V}\underline{C}^{-1}$. It can

be shown that

$$\Sigma_* = C_{22 \cdot 1}^{-1} GVG' C_{22 \cdot 1}^{-1}.$$

The following remark is immediate.

REMARK 3.1. Let $(A_1) - (A_9)$ hold. Then as $n \rightarrow \infty$,

$$S_1 = 12n\hat{\gamma}^2 \hat{\beta}_2' \Sigma_*^{-1} \hat{\beta}_2 \Big|_{(\hat{\beta}_1, \Delta_2/\sqrt{n})} \xrightarrow{L} \chi^2(p_2, \delta_1),$$

where the noncentrality parameter is $\delta_1 = 12\gamma^2 \Delta_2' C_{22 \cdot 1}^{-1} (GVG')^{-1} C_{22 \cdot 1}^{-1} \Delta_2$ and $\hat{\gamma}$ is a consistent estimate of $\gamma = \int f^2$.

Note that the distributions of $\hat{\beta}_2$ and S_1 do not depend on the nuisance parameter $\hat{\beta}_1$. The distribution of S_1 is asymptotically central chi-square when H_0 holds ($\Delta_2 = 0$) and noncentral chi-square under local alternatives $\hat{\beta}_2 = \Delta_2/\sqrt{n}$. The test of approximate level α is to reject H_0 if $S_1 > \chi_{\alpha, p_2}^2$.

In case that

$$(3.2) \quad GVG' = C_{22 \cdot 1}$$

there is further simplification in S_1 and δ_1 (also see Remark 4.2).

4. DROP IN DISPERSION TEST

McKean and Hettmansperger (1976, 1978b) proposed a test of H_0 based on the drop in a rank dispersion function between the full model and the reduced model. This data fitting criterion is appealing and is directly analogous to the least-squares test statistic. In this section the drop in dispersion statistic is considered for the weighted dispersion function (2.1).

Under H_0 , the reduced model of (3.1) contains the parameter β_1 . Let $\hat{\beta}_{1R}$ denote the reduced model estimate of β_1 obtained by minimizing the dispersion function $D(\underline{b}_1, 0)$ of (2.1) with respect to the p_1 variables in \underline{b}_1 . As before, $\hat{\beta}$ minimizes the dispersion function for the full model. The drop in dispersion test statistic is given by

$$S_2 = (12\hat{\gamma}/n)[D(\hat{\beta}_{1R}, 0) - D(\hat{\beta})],$$

where $\hat{\gamma}$ is a consistent estimate of γ .

The following remark concerns the asymptotic distribution of S_2 for testing purposes. The proof will be delayed to the Appendix.

REMARK 4.1. Let $(A_1) - (A_9)$ hold. Then for $\beta = (0, \Delta_2/\sqrt{n})$, as $n \rightarrow \infty$,

$$S_2 - 12\underline{U}(0)' \underline{G}' \underline{C}_{22}^{-1} \underline{G} \underline{U}(0) \xrightarrow{P} 0.$$

Using Remark 2.1 when $\beta = (0, \Delta_2/\sqrt{n})$ for the distribution of $\underline{U}(0)$, it follows that S_2 will be asymptotically chi-square if $\underline{G} \underline{V} \underline{G}' = \underline{C}_{22} \cdot 1$.

Note that the distribution of S_2 is free of β_1 since $S_2 | (\beta_1, \beta_2) \stackrel{L}{=} S_2 | (0, \beta_2)$ by a translation property of $\hat{\beta}$ and $\hat{\beta}_{1R}$. The following remark summarizes.

REMARK 4.2. Let $(A_1) - (A_9)$ hold. If $GVG' = C_{22.1}$, then as $n \rightarrow \infty$,

$$S_2 | (\beta_1, \Delta_2/\sqrt{n}) \xrightarrow{L} \chi^2(p_2, \delta_2),$$

where $\delta_2 = 12\gamma^2 \Delta_2' C_{22.1}^{-1} \Delta_2$.

The distribution of S_2 is asymptotically central chi-square when H_0 holds ($\Delta_2 = 0$) and noncentral chi-square under local alternatives $\beta_2 = \Delta_2/\sqrt{n}$. The test of approximate level α is to reject H_0 if $S_2 > \chi_{\alpha, p_2}^2$.

5. AN ALIGNED RANK TEST

Rank tests based on aligned ranks have been discussed by many authors for linear model problems, see Lehmann (1963), Adichie (1978), Sen and Puri (1977). The basic principle in aligning the observations is to estimate the nuisance parameters and to test the remaining parameters with a suitable statistic as if there were no nuisance parameters present. The resulting test typically has good large sample properties. Small sample results are difficult to obtain in general.

In the present context the aligned rank method requires a reduced model estimate of β_1 and a statistic to measure the relationship of the reduced model residuals to X_2 . Let $\hat{\beta}_{1R}$ be the reduced model estimate of β_1 of the previous section. Let

$$\hat{U} = U(\hat{\beta}_{1R}, 0) = (1/2)n^{-3/2} \underline{X}' \hat{B},$$

where $\hat{B} = (\hat{B}_1)$ using (2.2), (2.6) and $\hat{B}_1 = \sum_j w_{1j} \text{sgn}(\hat{Z}_1 - \hat{Z}_j)$, where $\hat{Z} = \underline{Y} - \underline{X}_1 \hat{\beta}_{1R} = (\hat{Z}_1)$ is the vector of reduced model residuals. Also form the partition

$$\hat{U} = \begin{pmatrix} \hat{U}_1 \\ \hat{U}_2 \end{pmatrix} = (1/2)n^{-3/2} \begin{pmatrix} \underline{X}'_1 \hat{B} \\ \underline{X}'_2 \hat{B} \end{pmatrix},$$

where \hat{U}_1 is $p_1 \times 1$ and \hat{U}_2 is $p_2 \times 1$.

The aligned rank test statistic will be a quadratic form in \hat{U}_2 . The following remark gives the necessary details. The proof is delayed to the Appendix.

REMARK 5.1. Let $(A_1) - (A_9)$ hold. Then as $n \rightarrow \infty$

$$\hat{U}_2 \Big|_{(\beta_1, \Delta_2/\sqrt{n})} \xrightarrow{L} N(\gamma C_{22 \cdot 1} \Delta_2, (1/12) GVG').$$

With this result it is easy to see that the appropriate quadratic form is

$$S_3 = 12 \hat{U}_2' (GVG')^{-1} \hat{U}_2.$$

The following remark is immediate from Remark 5.1.

REMARK 5.2. Let $(A_1) - (A_9)$ hold. Then as $n \rightarrow \infty$,

$$S_3 \Big|_{(\beta_1, \Delta_2/\sqrt{n})} \xrightarrow{L} \chi^2(p_2, \delta_3),$$

where the noncentrality parameter $\delta_3 = 12 \gamma^2 \Delta_2' C_{22 \cdot 1} (GVG')^{-1} C_{22 \cdot 1} \Delta_2$.

This is the same noncentrality parameter as in Remark 3.1 and if (3.2) holds it equals the noncentrality parameter of Remark 4.2.

The distribution of S_3 is asymptotically central chi-square under $H_0(\Delta_2 = 0)$ and noncentral chi-square under local alternatives $\beta_2 = \Delta_2/\sqrt{n}$. The test of approximate level α is to reject H_0 if $S_3 > \chi_{\alpha, p_2}^2$.

6. CHOICE OF WEIGHTS

In the unweighted case, $w_{ij} \equiv 1$, the asymptotic covariance matrix of the estimate $\hat{\beta}$ is a constant multiple of

$$(6.1) \quad \underline{C}^{-1} \underline{V} \underline{C}^{-1} = \underline{\Sigma}^{-1}.$$

As noted in Sievers (1983), this is a case of highest efficiency and the use of weights cannot improve on this. The finite sample size version of (6.1) is

$$(6.2) \quad (\underline{X}'\underline{WX})^{-1} \underline{X}'\underline{W}\underline{WX}(\underline{X}'\underline{WX})^{-1} = (\underline{X}'\underline{X})^{-1}$$

and when this condition holds there will be no loss of efficiency in using weights. A weight matrix will satisfy (6.2) if and only if there is a nonsingular matrix \underline{H} such that

$$(6.3) \quad \underline{WX} = \underline{XH}.$$

A goal in selecting weights would then be to satisfy (6.2) or (6.3). In the remainder of this section and in the next section some situations will be discussed where it is possible to select weights to retain efficiency and gain in other aspects.

The introduction indicated several one-way and higher order analysis of variance problems where restricted rankings are used. Such restricted comparison methods can be handled in the present context by using weights $w_{ij} = 1$ if the i th and j th observations are to be compared and zero otherwise. Examples in the next section indicate there need not be a loss of efficiency. Quade (1979b) and Silva and

Quade (1980) explored the use of weights proportional to a measure of within block variability for complete block designs.

In analysis of variance problems arising in practice a frequently occurring difficulty is that the assumptions for the basic additive model are in doubt. In such situations the use of restricted comparisons can be helpful. The additive model formally allows only shifts in location between groups but in practical applications there is often a drift from this that increases as the groups are further apart.

Neighboring groups in a design can be reasonably close in variation and shape of distribution due to similar experimental influences but this may deteriorate for groups that are more distant. The treatment may be affecting more than just the location of a distribution. Transformations can be tried to diminish this type of effect but they are not always successful and they introduce problems in interpreting the results. In these situations the comparisons between distant groups can be inappropriate. By comparing only neighboring groups the effects of group differences not formally specified in the model can be diminished. The focus is directly on the simple effects.

The specification of neighboring groups may be uncertain in a given problem but for simplicity it should be enough to compare observations only when they are in immediately adjacent groups in the design. This approach is discussed more in the examples of the next section. Modifications could be made to compare groups that differ by more than one level. The idea here is directly analogous to the comparisons of "relevant pairs" as proposed by Quade (1979a) for a multiple regression problem. He also suggested restricted comparisons for factorial designs.

7. EXAMPLES

Consider an analysis of variance model where data is available in $p + 1$ groups, labeled G_0, G_1, \dots, G_p , and the number of observations in the groups are denoted n_0, n_1, \dots, n_p , respectively. Let $n = n_0 + n_1 + \dots + n_p$. Suppose parameters are defined so that $y_i = \beta_0 + \beta_j + e_i$ if the i th observation is in group G_j . This is a convenient representation, similar to the means model except that G_0 is used as a reference group. The work of this paper is basically invariant with respect to reparametrizations and results from this simple model will carry over to other choices for defining parameters. Thus one-way and higher order factorial designs, block designs, etc. can be discussed in this framework.

With $\underline{\beta}' = (\beta_1, \dots, \beta_p)$, the $n \times p$ design matrix is

$$(7.1) \quad \underline{X} = \begin{pmatrix} \underline{0} & \underline{0} & \dots & \underline{0} \\ \underline{1}_1 & \underline{0} & \dots & \underline{0} \\ \underline{0} & \underline{1}_2 & \dots & \underline{0} \\ \underline{0} & \underline{0} & \dots & \underline{1}_p \end{pmatrix},$$

where $\underline{1}_j$ is an $n_j \times 1$ vector of "ones", $1 \leq j \leq p$, and $\underline{0}$ represents a vector of "zeros" of appropriate length. This design matrix \underline{X} should be centered to match the previous use of this symbol.

Let the weights depend only on group membership with $b_{jk} = b_{kj}$ the weight for comparing an observation in G_j with an observation in G_k .

The resulting weight matrix is

$$(7.2) \quad W = \begin{pmatrix} c_0 I_0 & -b_{01} J_{01} \dots -b_{0p} J_{0p} \\ -b_{10} J_{10} & c_1 I_1 \dots -b_{1p} J_{1p} \\ \vdots & \\ -b_{p0} J_{p0} & -b_{p1} J_{p1} \dots c_p I_p \end{pmatrix},$$

where I_j is an $n_j \times n_j$ identity matrix, $0 \leq j \leq p$, J_{kl} is an $n_k \times n_l$ matrix of "ones" and $c_j = \sum_{\substack{k=0 \\ k \neq j}}^p b_{jk} n_k$, $0 \leq j \leq p$.

Then

$$WX = \begin{pmatrix} -b_{01} n_{1-0} & -b_{02} n_{2-0} \dots -b_{0p} n_{p-0} \\ c_1 1_{-1} & -b_{12} n_{2-1} \\ -b_{21} n_{1-2} & c_2 1_{-2} \\ \vdots & \\ -b_{p1} n_{1-p} & -b_{p2} n_{2-p} \dots c_p 1_{-p} \end{pmatrix}$$

and

$$(7.3) \quad \underline{X}'\underline{W}\underline{X} = \begin{pmatrix} c_1 n_1 & -b_{12} n_1 n_2 & \dots & -b_{1p} n_1 n_p \\ -b_{21} n_2 n_1 & c_2 n_2 & \dots & -b_{2p} n_2 n_p \\ \vdots & & & \\ -b_{p1} n_p n_1 & & \dots & c_p n_p \end{pmatrix}$$

A matrix \underline{H} satisfying (6.3) is

$$(7.4) \quad \underline{H} = \begin{pmatrix} c_1 + b_{01} n_1 & (b_{02} - b_{12}) n_2 & \dots & (b_{0p} - b_{1p}) n_p \\ (b_{01} - b_{21}) n_1 & c_2 + b_{02} n_2 & & (b_{0p} - b_{2p}) n_p \\ \vdots & \vdots & & \vdots \\ (b_{01} - b_{p1}) n_1 & (b_{02} - b_{p2}) n_2 & & c_p + b_{0p} n_p \end{pmatrix}$$

and it remains to check if this matrix is nonsingular to verify no loss in efficiency. By premultiplying (6.3) by \underline{X}' it is equivalent, and sometimes easier, to verify that $\underline{X}'\underline{W}\underline{X}$ is nonsingular.

In this context the vector $\underline{U}(b)$ of (2.6) can be expressed more directly in terms of the comparisons between groups. Direct multiplication shows that the k^{th} element of $\underline{U}(b)$ is

$$(7.5) \quad U_k(b) = (1/2)n^{-3/2} \sum_{\substack{j=0 \\ j \neq k}}^p b_{jk} T_{jk},$$

where $T_{jk} = 2(\#G_j < G_k) - n_j n_k$ and $(\#G_j < G_k)$ is a Mann-Whitney statistic comparing groups G_j and G_k .

Several Treatments Versus a Control Suppose G_0 is a control group. With group weights $b_{0j} = 1$ for $j = 1, \dots, p$ and zero otherwise,

the treatment groups are compared to the control group but not to each other. The matrix H of (7.4) is nonsingular and there is no loss of efficiency. Remarks 2.1 and 2.3 along with (7.1)-(7.4) yield familiar results for this problem. Unequal sample sizes are readily handled. If covariates are present in the problem the design matrix can be modified accordingly and the tests of sections 4 or 5 can be used.

One-Way Analysis of Variance Suppose G_0, \dots, G_p represent $p + 1$ groups to be compared. In the unweighted case, $b_{jk} \equiv 1$, the quadratic form in $\underline{U}(0)$ based on Remark 2.1 yields the familiar Kruskal-Wallis test. As indicated in section 6, in some circumstances the use of restricted comparisons may be beneficial.

Suppose there is a natural order in the treatment groups; for instance the groups may correspond to increasing dose level, to increasing age of subjects or to geographic locations along a path. The discussion in section 6 suggests that the use of adjacent comparisons can be beneficial in such cases. Let $b_{jk} = 1$ if $j = k - 1$ for $k = 1, \dots, p$ and zero otherwise. Then observations are compared only to observations in an immediately adjacent group. An examination of (7.3) shows it is nonsingular, so H is nonsingular and there is no loss of efficiency. Thus the use of adjacent comparisons can be a useful alternative to the unweighted case, especially in the presence of model deficiencies.

Consider yet another pattern of weights for this problem. Let the weight of a comparison between groups G_j and G_k be given by $b_{jk} = b_{kj} = 1/n_j n_k$. The effect is to give less weight to observations

in groups with larger sample sizes. With these weights, (7.3) simplifies considerably and it is easy to show \underline{H} is nonsingular. Thus there is no loss of efficiency. Sievers (1983) discussed an influence matrix to measure the influence of the i^{th} observation on the estimate of β_j . This $n \times p$ matrix for the estimate $\hat{\beta}$ minimizing (2.1) is $\underline{WX}(X'WX)^{-1}$. The weights here yield the same influence matrix as the unweighted case. This follows by direct computation.

Ordered Alternatives In the notation of the previous one-way layout the ordered alternative specifies $0 \leq \beta_1 \leq \dots \leq \beta_p$ with some strict inequality. Making all comparisons requires that G_j and G_k be compared for all $j < k$. Test statistics based on ranks $\sum_{j < k} T_{jk}$ and $\sum_{j < k} T_{jk}/n_j n_k$ have been proposed for this problem, see Jonckheere (1954). These statistics are linear combinations of $\underline{U}(0)$ as given in (7.5) using suitable coefficients and weights b_{jk} . Tryon and Hettmansperger (1973) showed value in using only adjacent comparisons with a statistic $\sum_{j=1}^P a_j T_{j-1,j}$. This can be obtained from (7.5) with weights $b_{jk} = 1$ if $j = k - 1$ and zero otherwise.

By choosing appropriate weights b_{jk} , these linear combination statistics can have a wide variety of coefficients. There is some value in this general view. Many familiar results on means, variances and asymptotic normality follow as special cases of Remark 2.1. Unbalanced cases cause no special trouble.

By using the aligned rank method of section 5 the presence of nuisance parameters or covariates can be handled in a straightforward, unified manner. The method can be adapted to provide a suitable test statistic for variations in the standard model. For instance, if the

groups are observed in incomplete blocks the use of suitable 0-1 weights can restrict comparisons to the within block comparisons to cancel out block effects and then a suitable linear combination of the resulting T_{jk} could be chosen to test the ordered alternative. As another example suppose the alternative hypothesis only specifies a partial ordering on the β_j instead of a complete ordering. For such a case the use of 0-1 weights can restrict comparisons to those matching the alternative hypothesis.

Two-Factor Analysis of Variance A 2 x 2 layout will be discussed for simplicity but extensions to larger tables and higher dimensions will be apparent. The notation of the beginning of this section will be retained, but with a change of parameters, to avoid introducing double subscripts. Label the cells and the model parameters as follows:

G_0	G_1
G_2	G_3

μ	$\mu + \beta_1$
$\mu + \alpha_1$	$\mu + \alpha_1 + \beta_1 + \gamma_1$

Thus α_1 is a row effect, β_1 a column effect and γ_1 an interaction effect. The parameters here are related to the parameters β of the beginning of this section by a nonsingular linear transformation and as a consequence results obtained in one case apply to the other.

Using the earlier notation for sample sizes, the $n \times 3$ design matrix is

$$X = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1_1 & 0 \\ 1_2 & 0 & 0 \\ 1_3 & 1_3 & 1_3 \end{pmatrix}$$

and it should be centered yet to conform to earlier notation.

The main weighting scheme to be considered here is the scheme that uses only within row/within column comparisons, abbreviated WR/WC. In this scheme groups are compared (nonzero weights) if they are in the same row or same column while diagonal comparisons are omitted with zero weights ($b_{03} = b_{12} = 0$). In this way comparisons are made between observations that differ only in the level of one factor. This follows the general rationale of section 6.

If nonzero weights are "one", the WR/WC weight matrix is

$$(7.6) \quad \tilde{W} = \begin{pmatrix} (n_2 + n_3)I_0 & -J_{01} & -J_{02} & 0 \\ -J_{10} & (n_1 + n_4)I_1 & 0 & -J_{13} \\ -J_{20} & 0 & (n_1 + n_4)I_2 & -J_{23} \\ 0 & -J_{31} & -J_{32} & (n_2 + n_3)I_3 \end{pmatrix} .$$

A straightforward calculation shows that (6.3) holds and there is no loss of efficiency with these weights.

The earlier work applied here shows that the weights $b_{jk} = 1/n_j n_k$ retain full efficiency. This is also the case if these weights are used for the nonzero weights in the WR/WC scheme.

Consider the model above with no interaction ($\gamma_1 = 0$). If all comparisons are used in testing $H_0: \alpha_1 = 0$ the column effect β_1 is a nuisance parameter and the methods of section 4 or 5 would be needed. However, if the WR/WC weight scheme (7.6) is used the nuisance parameter cancels out and the test statistic can avoid estimating it. This is a desirable feature and it clearly extends to larger sized layouts. There is no loss of efficiency if the sample sizes are equal. However, with

unbalanced cases there can be a loss. Perhaps (7.6) can be further modified to resolve this problem.

When the two-way layout is larger in size than 2×2 there is another possible weighting scheme consistent with the remarks of section 6. In the WR/WC scheme, instead of comparing a group to all other groups in the same row or column, compare each group to only its immediate neighbors in the same row and column. This adjacent WR/WC plan should prove to be useful and warrants further study.

In the model with no interaction consider testing the hypothesis of no column effect, $H_0: \beta_1 = 0$, with the aligned rank procedure of section 5. To align the rows, the estimate of α_1 suggested in section 5 is a rank estimate. The literature is quite varied on this point and other \sqrt{n} -consistent estimates have been suggested, for example, differences of row means or medians. Suppose the estimate of α_1 is denoted by a_1 and the cells are aligned by subtracting a_1 from all observations in the second row. The aligned rank statistic \hat{U}_2 depends on a_1 and to examine this relationship the derivative $d\hat{U}_2/da_1$ will be calculated. It will be shown that this derivative is zero for certain choices of weights. Thus the use of appropriate weights can reduce the effect of alignment on the test statistic.

First note that \hat{U}_2 is the second element of $\hat{X}'\hat{B}$ and except for a constant is given by

$$(7.7) \quad \hat{U}_2 = b_{01}T_{01} + b_{21}T_{21} + b_{03}T_{03} + b_{23}T_{23}$$

where $T_{jk} = 2n_j n_k (\int F_j dF_k - (1/2))$ and F_j is the empirical cdf of the (aligned) observations in group G_j , $j = 0, 1, 2, 3$.

To avoid complications without losing the main point, consider the F_j as continuous cdfs with densities. Without loss of generality assume the true value of α_1 is zero. Under H_0 , $\beta_1 = 0$. Then write $F_0(x) = F(x)$, $F_1(x) = F(x)$, $F_2(x) = F(x+a_1)$ and $F_3(x) = F(x+a_1)$. Substituting these into (7.7), differentiating and evaluating at $a_1 = 0$ yields

$$\hat{d}U_2/da_1 \Big|_{a_1=0} = 2(n_1n_2b_{12} - n_0n_3b_{03})/f^2$$

This derivative is zero when $b_{12} = b_{03} = 0$, that is, in the case of WR/WC comparisons. It is also zero when $b_{jk} = 1/n_jn_k$. In these two cases, at least, the rate of change of \hat{U}_2 with respect to the aligning quantity a_1 is zero.

Block Designs The weighting schemes can be used with block designs. Some care must be taken, however, since the asymptotic results here apply when the within-cell sample sizes grow large and not, for example, in a case of a fixed block size with the number of blocks growing large. Assumptions $(A_1) - (A_3)$ may not hold. The asymptotic results here can also apply when a given design is replicated and the number of replications grows large.

Consider a case where a basic design with m observations per cell, m large, is replicated with replications corresponding to blocks. By using zero weights the between block comparisons can be eliminated and block effects would cancel out. Tests of hypotheses about the parameters in the basic design (or about a subset of them) can be constructed from the general results here. Covariates can be handled in the same framework. Unbalanced sample sizes cause no special problem.

8. APPENDIX

Proof of Remark 4.1. Some additional notation is needed. Let $D^*(\underline{\Delta}) = (1/n)D(\underline{\Delta}/\sqrt{n})$. Then $\hat{\underline{\Delta}} = \sqrt{n}\hat{\underline{\beta}}$ minimizes $D^*(\underline{\Delta})$. Similarly, $\hat{\underline{\Delta}}_{1R} = \sqrt{n}\hat{\underline{\beta}}_{1R}$ minimizes $D^*(\underline{\Delta}_1, 0)$. Define a quadratic function $Q(\underline{\Delta}) = \gamma \underline{\Delta}' \underline{C} \underline{\Delta} - 2 \underline{\Delta}' \underline{U}(0) + D^*(0)$. Then $\underline{\Delta}^* = \gamma^{-1} \underline{C}^{-1} \underline{U}(0)$ minimizes $Q(\underline{\Delta})$ and $\underline{\Delta}_{1R}^* = \gamma^{-1} \underline{C}_{11}^{-1} \underline{U}_1(0)$ minimizes $Q(\underline{\Delta}_1, 0)$, where $\underline{U}_1(0)$ is the first p_1 elements of $\underline{U}(0)$.

It is sufficient to use γ in place of $\hat{\gamma}$ in S_2 . Then

$$\begin{aligned} S_2 &= (12\gamma)[D^*(\hat{\underline{\Delta}}_{1R}, 0) - D^*(\hat{\underline{\Delta}})] \\ &= (12\gamma)\{[D^*(\hat{\underline{\Delta}}_{1R}, 0) - Q(\hat{\underline{\Delta}}_{1R}, 0)] + [Q(\hat{\underline{\Delta}}_{1R}, 0) - Q(\underline{\Delta}_{1R}^*, 0)] \\ &\quad + [Q(\underline{\Delta}_{1R}^*, 0) - Q(\underline{\Delta}^*)] + [Q(\underline{\Delta}^*) - Q(\hat{\underline{\Delta}})] \\ &\quad + [Q(\hat{\underline{\Delta}}) - D^*(\hat{\underline{\Delta}})]\}. \end{aligned}$$

Now Theorem 6.1 and Lemma 6.4 of Sievers (1983) can be used to show that if $\underline{\beta} = 0$, the terms above, except the middle one, converge in probability to zero. This can be extended to the case $\underline{\beta} = (0, \underline{\Delta}_2/\sqrt{n})$ by contiguity. But the middle term above is $12[\underline{U}(0)' \underline{C}^{-1} \underline{U}(0) - \underline{U}_1(0)' \underline{C}_{11}^{-1} \underline{U}_1(0)] = 12 \underline{U}(0)' \underline{G}' \underline{C}_{22 \cdot 1}^{-1} \underline{G} \underline{U}(0)$ and the result follows.

Proof of Remark 5.1. By a translation property of the reduced model

estimate $\hat{\underline{\beta}}_{1R}$ is enough to prove the result when $\underline{\beta}_1 = 0$ since the distribution of $\hat{\underline{U}}_2$ is unaffected by the value of $\underline{\beta}_1$.

From Remark 2.3 it follows that $\hat{\underline{\Delta}}_1 = \sqrt{n}\hat{\underline{\beta}}_{1R} \Big|_0$ has a limiting normal distribution and is $O_p(1)$. This, with Remark 2.2, implies

$$\underline{U}(\hat{\beta}_{1R}, 0) - \underline{U}(0) + \gamma \underline{C} \begin{pmatrix} \hat{\Delta}_1 \\ 0 \end{pmatrix} \Big|_0 \xrightarrow{P} 0.$$

Using a contiguity argument,

$$\hat{\underline{U}} - \underline{U}(0) + \gamma \begin{pmatrix} C_{11} \\ C_{21} \end{pmatrix} \hat{\Delta}_1 \Big|_{(0, \Delta_2/\sqrt{n})} \xrightarrow{P} 0.$$

This result continues to hold if the fixed matrix \underline{G} is multiplied on the left and this drops out the term containing $\hat{\Delta}_1$. Thus

$$\underline{G}\hat{\underline{U}} - \underline{G}\underline{U}(0) \Big|_{(0, \Delta_2/\sqrt{n})} \xrightarrow{P} 0.$$

But by Remark 2.1,

$$\underline{G}\underline{U}(0) \Big|_{(0, \Delta_2/\sqrt{n})} \xrightarrow{L} N(\gamma C_{22}^{-1} \Delta_2, (1/12) \underline{G}\underline{V}\underline{G}')$$

and so $\underline{G}\hat{\underline{U}}$ has this same limiting distribution. The proof is finished by noting that $\underline{G}\hat{\underline{U}} = \hat{\underline{U}}_2 - C_{21} C_{11}^{-1} \hat{\underline{U}}_1 \doteq \hat{\underline{U}}_2$ since $\hat{\underline{U}}_1 \doteq 0$, being the derivative (essentially) of the reduced model dispersion function evaluated at the reduced model estimate.

REFERENCES

- Adichie, J.N. (1978), "Rank Tests of Sub-Hypotheses in the General Linear Regression," *Annals of Statistics*, 6, 1012-1026.
- Hettmansperger, T.P. and McKean, J.W. (1978a), "Statistical Inference Based on Ranks," *Psychometrika*, 43, 69-79.
- Hodges, J.L., Jr. and Lehmann, E.L. (1962), "Rank Methods for Combination of Independent Experiments in the Analysis of Variance," *Annals of Mathematical Statistics*, 33, 482-497.
- Jonckheere, A.R. (1954), "A Distribution Free K - Sample Test Against Ordered Alternatives," *Biometrika*, 41, 133-145.
- McKean, J.W. and Hettmansperger, T.P. (1976), "Tests of Hypotheses Based on Ranks in the General Linear Model," *Communications in Statistics - Theory and Methods - A*, 8, 693-709.
- McKean, J.W. and Hettmansperger, T.P. (1978b), "A Robust Analysis of the General Linear Model Based on One Step R - Estimates" *Biometrika*, 65, 571-579.
- Quade, D. (1979a), "Regression Analysis Based On the Signs of the Residuals," *Journal of the American Statistical Association*, 74, 411-417.
- Quade, D. (1979b), "Using Weighted Rankings in the Analysis of Complete Blocks With Additive Block Effects," *Journal of the American Statistical Association*, 74, 680-683.
- Sen, P.K. and Puri, M.L. (1977), "Asymptotically Distribution-Free Aligned Rank Order Tests for Composite Hypotheses for General Multivariate Linear Models," *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, 39, 175-186.
- Sievers, G.L. (1983), "A Weighted Dispersion Function for Estimation in Linear Models," *Communications in Statistics - Theory and Methods - A*, 12, 1161-1179.
- Silva, C. and Quade, D. (1980), "Evaluation of Weighted Rankings Using Expected Significance Level," *Communications in Statistics - Theory and Methods - A*, 9, 1087-1096.
- Tryon, P.V. and Hettmansperger, T.P. (1973), "A Class of Non-parametric Tests for Homogeneity Against Ordered Alternatives," *Annals of Statistics*, 1, 1061-1070.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Technical Report #71	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Testing Hypotheses in Linear Models with Weighted Rank Statistics		5. TYPE OF REPORT & PERIOD COVERED Technical Report 1983
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Gerald L. Sievers		8. CONTRACT OR GRANT NUMBER(s) N 00014-78-C-0637
9. PERFORMING ORGANIZATION NAME AND ADDRESS Western Michigan University Kalamazoo, Michigan 49008		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 042-407
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics and Probability Program		12. REPORT DATE November 1983
		13. NUMBER OF PAGES 30
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Analysis of variance, dispersion function, robust tests, aligned ranks, rank estimates, weights		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Tests of hypotheses for the parameters in a general linear model are considered based on weighted rank statistics. Results are presented for tests based on a rank estimate, tests based on drop in dispersion and aligned rank tests. Weights can be used to focus the analysis on simple effects and provide an additional degree of robustness to rank tests. Several analysis of variance applications are discussed.		

DD FORM 1473
1 JAN 73

EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)