Analysis of Nonadditive Multiway Classifications

ROBERT J. BOIK and MERVYN G. MARASINGHE*

This article considers the problems of testing additivity and estimating σ^2 in unreplicated multiway classifications. To model nonadditivity and jointly estimate σ^2, the interaction parameter space must be restricted; otherwise the model is saturated. The parameterization we use is a multiway extension of the two-way multiplicative interaction model of Mandel (1971) and Johnson and Graybill (1972a). For example, in a three-way classification, we model interaction as $\theta_{ijk} = \lambda\delta_{1i}\delta_{2j}\delta_{3k}$. This structure is a special case of the k-mode principal components model, which has received considerable attention in the psychometric literature (Kapteyn, Neudecker, and Wansbeek 1986). We construct an exact test of $\lambda = 0$ and propose an estimator of σ^2 that can be used when interaction has been detected. Our test is an approximation to the likelihood ratio test (LRT) of H_o : $\lambda = 0$. The proposed test has essentially the same power as the LRT but is easier to compute, and the exact null distribution of the test statistic is known. Selected percentiles of the null distribution are given for three-way classifications. For large $|\lambda/\sigma|$, a transformation of the test statistic is shown to be approximately distributed as a noncentral F and can be used to compute the power of the test. The test and estimator are illustrated on a data set having three rows, three columns, and four layers.

KEY WORDS: ANOVA; Interaction; Likelihood ratio test; Multiplicative model; Principal components; Reduced-rank model.

1. INTRODUCTION

This article develops inference procedures for the highest-order interaction in unreplicated multiway classifications. The procedures require that exactly one observation be made under each treatment combination. The procedures could be extended to equally or proportionally replicated models, but that is not the purpose of this article. Unfortunately, the procedures do not readily generalize to incomplete factorial arrangements, even when cells are empty by design.

To achieve generality with respect to the number of classifications, we depart, somewhat, from the usual notation. The saturated fixed effects model for a k-way classification having one observation per treatment combination can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\theta} + \boldsymbol{\epsilon}, \qquad (1.1)$$

where $\mathbf{y}$ is an n-vector of random variables, $n = \Pi_{i=1}^{k} m_i$; m_i is the number of levels of the ith classification; $\mathbf{X}$: $n \times p$ is the usual design matrix coding for all main effects and all interactions except for the k-way interaction, $p = \Pi_{i=1}^{k}(m_i + 1) - n$; $\boldsymbol{\beta}$ is the corresponding p-vector of main effect and interaction parameters; $\boldsymbol{\theta}$ is an n-vector of k-way interaction parameters; and $\boldsymbol{\epsilon} \sim N(\mathbf{O}, \sigma^2\mathbf{I}_n)$. Each element in $\mathbf{y}$ is associated with a unique k-tuple of indices, the ith index running from 1 to m_i. The elements in $\mathbf{y}$ are assumed to be ordered such that $m_1 \leq \cdots \leq m_k$, the first index changing slowest, the kth index changing fastest, etcetera.

Typically, σ^2 is estimated by dropping $\boldsymbol{\theta}$ and fitting the reduced additive model. If k-way interaction exists, however, the resulting estimator is distributed as a multiple of a noncentral chi squared and the usual tests of main effects and lower-order interactions are conservative. To check for nonadditivity, a test of $H_o : E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ against $H_a : E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\theta}$ would be useful.

The main difficulty in testing additivity is that, without restricting the interaction parameter space, the full model is saturated. No degrees of freedom are available for estimating σ^2, so the usual F test cannot be done. We restrict the parameter space by adopting a multiplicative model for k-way interaction. The model is a generalization of the two-way multiplicative model suggested by Mandel (1971) and Johnson and Graybill (1972a). The Johnson–Graybill test of additivity in nonreplicated two-way classifications is briefly reviewed in Section 2.1. A thorough review was given by Milliken and Johnson (1989). In Section 2.2, the k-way generalization is described and the associated likelihood ratio test (LRT) is given. The null distribution of the LRT statistic, however, is unknown.

The new results are presented in Section 3, where we introduce a statistic that bounds the LRT statistic from below. The exact null distribution of the bounding statistic is found, and a table of upper percentiles when $k = 3$ is given. Under H_a, we show that the LRT and lower bound statistics are each, approximately, distributed as the same noncentral F with noncentrality parameter $\|\boldsymbol{\theta}/\sigma\|^2$. Thus the lower bound statistic is a sensible criterion for testing additivity. Section 3 also gives an estimator of experimental error that can be used when k-way interaction is detected. The estimator is based on the distribution of the residual sum of squares after fitting the multiplicative model, assuming large $\|\boldsymbol{\theta}/\sigma\|$. The new methods are illustrated in Section 4.

2. CURRENT TESTS FOR INTERACTION IN MULTIWAY CLASSIFICATIONS

2.1 The Johnson–Graybill Test for Two-Way Interaction

The Johnson–Graybill–Mandel model for nonreplicated two-way classifications is (1.1) where $k = 2$,

$$\boldsymbol{\theta} = (\boldsymbol{\delta}_1 \otimes \boldsymbol{\delta}_2)\lambda, \qquad (2.1)$$

λ is a scalar constant, and the vectors $\boldsymbol{\delta}_1 : m_1 \times 1$ and $\boldsymbol{\delta}_2 : m_2 \times 1$ are each constrained to have unit norm and

* Robert J. Boik is Associate Professor, Department of Mathematical Sciences, Montana State University, Bozeman, MT 59717. Mervyn G. Marasinghe is Associate Professor, Department of Statistics, Iowa State University, Ames, IA 50011. A portion of the second author's work was supported by a grant from the Iowa State University Achievement Foundation.

Journal of the American Statistical Association
December 1989, Vol. 84, No. 408, Theory and Methods

to sum to 0. The elements of $\boldsymbol{\theta}$ can also be written as $\theta_{ij} = \lambda\delta_{1i}\delta_{2j}$. Johnson and Graybill (1972a) showed that the maximum likelihood estimator (MLE) of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\delta}}_1 \otimes \hat{\boldsymbol{\delta}}_2)\hat{\lambda},$$

where $\hat{\boldsymbol{\delta}}_1 = \mathbf{v}_1(\mathbf{E}'\mathbf{E})$; $\hat{\boldsymbol{\delta}}_2 = \mathbf{v}_1(\mathbf{E}\mathbf{E}')$; $\hat{\lambda} = (\hat{\boldsymbol{\delta}}_1 \otimes \hat{\boldsymbol{\delta}}_2)'\mathbf{e}$; $\mathbf{e}$ is the n-vector of residuals from the fitted additive model, $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$, $\text{vec}(\mathbf{E}) = \mathbf{e}$ where $\mathbf{E}$ is an $m_2 \times m_1$ matrix; and $\mathbf{v}_1(\mathbf{A})$ is the unit norm characteristic vector associated with the maximum characteristic root of $\mathbf{A}$. The LRT statistic for testing $H_o : \lambda = 0$ in Model (1.1) with (2.1) is

$$w_2 = \hat{\lambda}^2/\mathbf{e}'\mathbf{e}, \tag{2.2}$$

where the subscript on w signals that it is two-way interaction being tested. The null distribution of w_2 was given by Johnson and Graybill as $w_2 \sim r_1(\mathbf{A})/\text{tr}(\mathbf{A})$, where $r_1(\mathbf{A})$ is the maximum characteristic root of $\mathbf{A}$, $\mathbf{A} \sim W_p(q, \mathbf{I})$, $p = m_1 - 1$, and $q = m_2 - 1$. Null percentiles of w_2 can be found in Schuurmann, Krishnaiah, and Chattopadhyay (1973), Krzanowski (1979), and Boik (1989a). Extensions to more than one multiplicative component in two-way classifications were described by Krishnaiah and Yochmowitz (1980).

2.2 The Likelihood Ratio Test for *k*-Way Interaction

The k-way generalization of the Mandel–Johnson–Graybill multiplicative structure (2.1) is

$$\boldsymbol{\theta} = \boldsymbol{\Delta}\lambda, \tag{2.3}$$

where $\boldsymbol{\Delta} = \boldsymbol{\delta}_1 \otimes \cdots \otimes \boldsymbol{\delta}_k$, each $\boldsymbol{\delta}_i : m_i \times 1$ is constrained to have unit norm and to sum to 0, and λ is a scalar constant. Multiplicative structures like (2.3) already have a sizable literature. For $k = 3$, (2.3) is a special case of Carroll and Chang's (1970) three-way singular value decomposition, which, in turn, is a special case of Tucker's (1966) three-mode principal component model. Tucker's model, extended to k-modes, is (2.3), in which each $\boldsymbol{\delta}_i$ may have more than one column, each $\boldsymbol{\delta}_i$ is semiorthogonal, and λ is a vector (see Kapteyn, Neudecker, and Wansbeek 1986; ten Berge, de Leeuw, and Kroonenberg 1987). Carvalho (1978) and Kettenring (1983) applied (2.3) to interaction in unreplicated three-way classifications. Research on the k-mode model has focused on obtaining point estimates of the parameters. Construction of tests and confidence intervals has received little attention.

The MLE's of σ^2 and θ in Model (1.1) with (2.3) are

$$\hat{\sigma}^2 = [\mathbf{e}'\mathbf{e} - \hat{\lambda}^2]/n \tag{2.4a}$$

and

$$\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\Delta}}\hat{\lambda}, \tag{2.4b}$$

where $\hat{\boldsymbol{\Delta}} = \hat{\boldsymbol{\delta}}_1 \otimes \cdots \otimes \hat{\boldsymbol{\delta}}_k$; $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$; $\hat{\boldsymbol{\beta}}$ is the ordinary least squares (OLS) estimator of $\boldsymbol{\beta}$ in the additive model;

$$\hat{\boldsymbol{\delta}}_i = \hat{\boldsymbol{\Delta}}_i'\mathbf{e} \times (\mathbf{e}'\hat{\boldsymbol{\Delta}}_i\hat{\boldsymbol{\Delta}}_i'\mathbf{e})^{-1/2}, \quad i = 1, \ldots, k; \tag{2.4c}$$

$$\hat{\boldsymbol{\Delta}}_i = \hat{\boldsymbol{\delta}}_1 \otimes \cdots \otimes \hat{\boldsymbol{\delta}}_{i-1} \otimes \mathbf{I}_{m_i} \otimes \hat{\boldsymbol{\delta}}_{i+1} \otimes \cdots \otimes \hat{\boldsymbol{\delta}}_k;$$

and

$$\hat{\lambda} = \hat{\boldsymbol{\Delta}}'\mathbf{e}. \tag{2.4d}$$

Note that $\hat{\boldsymbol{\Delta}}_i$ is not a partition of $\hat{\boldsymbol{\Delta}} : \hat{\boldsymbol{\Delta}}$ is $n \times 1$ and $\hat{\boldsymbol{\Delta}}_i$ is $n \times m_i$. Equation (2.4c) can be solved by following Kapteyn et al. (1986).

Under $H_o : \lambda = 0$, the MLE of σ^2 is $\mathbf{e}'\mathbf{e}/n$. Thus the LRT can be given as follows: Reject H_o for large values of

$$w_k = \hat{\lambda}^2/\mathbf{e}'\mathbf{e}. \tag{2.5}$$

The numerator of (2.5) is the maximal sum of squares of a residual contrast having a coefficient vector structured as $\boldsymbol{\Delta}$ in (2.3). Thus the LRT statistic is the proportion of the residual sum of squares attributable to the maximal k-way product contrast.

If $n = 2^{k-2}m_{k-1}m_k$, then $w_k \sim r_1(\mathbf{A})/\text{tr}(\mathbf{A})$, where $\mathbf{A} \sim W_p(q, \mathbf{I}, \boldsymbol{\Omega})$, $p = m_{k-1} - 1$, $q = m_k - 1$, and the single nonzero root of $\boldsymbol{\Omega}$ is $\hat{\lambda}^2/\sigma^2$. Otherwise, the distribution of w_k is unknown. Farmer (1987) used simulation to estimate the null moments and approximated the distribution of w_3 by a beta distribution.

2.3 A Score Test for *k*-Way Interaction

A score test for interaction in three-way classifications was proposed by Harter and Lum (1962). Their test is a three-way extension of Tukey's (1949) 1 df for nonadditivity. The k-way extension is straightforward. The underlying model assumes that $\boldsymbol{\theta}$ can be written as in (2.3), in which $\boldsymbol{\delta}_i$ is proportional to the m_i-vector of main effects for the ith classification. To test $H_o : \lambda = 0$, main effects and lower-order interactions are estimated assuming an additive model and imposing the usual sum to 0 restrictions. Estimates of $\boldsymbol{\Delta}$ and λ are then obtained by setting each $\boldsymbol{\delta}_i$ equal to the corresponding vector of estimated main effects and employing OLS to estimate λ. Using a conditional argument (Milliken and Graybill 1970), it is readily shown that, under H_o, $F = [\text{SS}_k/(\mathbf{e}'\mathbf{e} - \text{SS}_k)]v$ has a central F distribution with 1 and $v = \Pi_{i=1}^{k}(m_i - 1) - 1$ df, where SS_k is a 1 df sum of squares for nonadditivity:

$$\text{SS}_k = (\hat{\boldsymbol{\Delta}}'\mathbf{e})^2/\hat{\boldsymbol{\Delta}}'\hat{\boldsymbol{\Delta}},$$

$\hat{\boldsymbol{\Delta}} = \hat{\boldsymbol{\delta}}_1 \otimes \cdots \otimes \hat{\boldsymbol{\delta}}_k$, and $\hat{\boldsymbol{\delta}}_i$ is the vector of estimated main effects for the ith classification. When $k = 2$, the score test reduces to Tukey's test.

3. A NEW TEST FOR INTERACTION IN MULTIWAY CLASSIFICATIONS

This section introduces an easily computed statistic, u_k, having a known null distribution. We argue that the test based on u_k approximates the LRT, and we give an associated estimator of σ^2 that can be used when $\lambda \neq 0$. The test statistic is based on estimators of $\{\boldsymbol{\delta}_i\}$, which are obtained sequentially. In the first step, the k-way classification is cast as a two-way $n/m_1 \times m_1$ classification and $\boldsymbol{\delta}_1$ is estimated as in the Johnson–Graybill test of Section 2.1. In the second step, $\boldsymbol{\delta}_1$ is set equal to its estimate and the dimension of the classification is reduced from k to k

− 1. The first step is then repeated and $\boldsymbol{\delta}_2$ is estimated, etcetera. The resulting test statistic is

$$u_k = \tilde{\lambda}^2/\mathbf{e}'\mathbf{e}, \qquad (3.1a)$$

where $\tilde{\lambda} = \tilde{\boldsymbol{\Delta}}'\mathbf{e}$, $\tilde{\boldsymbol{\Delta}} = \tilde{\boldsymbol{\delta}}_1 \otimes \cdots \otimes \tilde{\boldsymbol{\delta}}_k$,

$$\tilde{\boldsymbol{\delta}}_i = \mathbf{v}_1(\mathbf{E}_i'\mathbf{E}_i), \qquad i = 1, \ldots, k, \qquad (3.1b)$$

$\mathbf{E}_i$ is $s_i \times m_i$, $\operatorname{vec}(\mathbf{E}_1) = \mathbf{e}$,

$$\operatorname{vec}(\mathbf{E}_i) = \mathbf{E}_{i-1}\tilde{\boldsymbol{\delta}}_{i-1}, \qquad i = 2, \ldots, k, \qquad (3.1c)$$

and

$$s_i = n \div \prod_{j=1}^{i} m_j, \qquad i = 1, \ldots, k. \qquad (3.1d)$$

If $n = 2^{k-2}m_{k-1}m_k$, then $u_k = w_k$. Otherwise, $u_k < w_k$, because $\tilde{\lambda}^2$ is the maximal sum of squares associated with a k-way product residual contrast. In particular, $u_2 = w_2$, the Johnson–Graybill statistic. A listing of the SAS IML code (SAS 1985) for computing u_k is available from the senior author.

By reordering the k indices, $k!/2$ lower bound statistics can be obtained. The order $m_1 \le \cdots \le m_k$ is recommended, because, given steps 1, . . . , $(i - 1)$, it minimizes the number of parameters estimated during the ith step. The result is a statistic that is closer, on average, to the LRT statistic than one obtained from an alternative order. For factors having the same number of levels, we suggest ordering the indices according to the magnitude of the corresponding main effect sums of squares (largest to smallest).

3.1 Null Distribution

From the definitions in (3.1), it is readily shown that u_k can be written as

$$u_k = \prod_{i=1}^{k-1} [r_1(\mathbf{E}_i\mathbf{E}_i')/\operatorname{tr}(\mathbf{E}_i\mathbf{E}_i')].$$

Furthermore, when H_o is true, theorem 8.1 in James (1954) can be used to show (a)

$$\operatorname{vec}(\mathbf{E}_i) \times [\operatorname{tr}(\mathbf{E}_i\mathbf{E}_i')]^{-1/2} \sim \left[\bigotimes_{j=i}^{k} (\mathbf{I} - m_j^{-1}\mathbf{J}_j)\right]\mathbf{z}_i \times \left\{\mathbf{z}_i'\left[\bigotimes_{j=i}^{k} (\mathbf{I} - m_j^{-1}\mathbf{J}_j)\right]\mathbf{z}_i\right\}^{-1/2}$$

for $i = 1, \ldots, (k - 1)$ where $\mathbf{z}_i$ has an invariant distribution on the Stiefel manifold $V(1, s_{i-1})$, $s_o = n$, and $\mathbf{J}_j : m_j \times m_j$ is a matrix of ones; and (b) the characteristic roots of $\mathbf{E}_i\mathbf{E}_i'$ are distributed independently of the characteristic vectors of $\mathbf{E}_i\mathbf{E}_i'$. Together, these results give the null distribution of u_k summarized in Theorem 1.

Theorem 1. Under $H_o : \lambda = 0$,

$$u_k \sim \prod_{i=1}^{k-1} l_i,$$

where $\{l_i\}$ are independently distributed as $l_i \sim r_1(\mathbf{A}_i)/\operatorname{tr}(\mathbf{A}_i)$, $\mathbf{A}_i \sim W_{p_i}(q_i, \mathbf{I})$, $p_i = m_i - 1$, and $q_i = \Pi_{j=i+1}^{k}(m_j - 1)$.

Define l_{ih} by

$$l_{ih} = \prod_{j=i}^{h} l_j.$$

Table 1. Upper Percentiles of the Null Distribution of u_3

			$1-\alpha$		
m_1	m_2	m_3	.90	.95	.99
3	3	3	.88906	.92342	.96673
3	3	4	.78987	.83249	.89848
3	3	5	.72123	.76426	.83660
3	3	6	.67162	.71312	.78600
3	3	7	.63399	.67353	.74485
3	3	8	.60433	.64191	.71093
3	3	9	.58025	.61600	.68251
3	3	10	.56022	.59432	.65835
3	4	4	.67356	.71699	.79266
3	4	5	.59992	.64042	.71447
3	4	6	.54894	.58629	.65635
3	4	7	.51130	.54584	.61165
3	4	8	.48219	.51431	.57617
3	4	9	.45889	.48895	.54726
3	4	10	.43974	.46803	.52319
3	5	5	.52607	.56252	.63150
3	5	6	.47601	.50890	.57232
3	5	7	.43955	.46952	.52795
3	5	8	.41171	.43921	.49335
3	5	9	.38992	.41508	.46553
3	5	10	.37274	.39548	.44260
3	6	6	.42724	.45651	.51381
3	6	7	.39255	.41849	.47061
3	6	8	.36797	.38997	.43728
3	6	9	.35114	.36902	.41075
3	6	10	.33964	.35430	.38942
4	4	4	.55439	.59384	.66733
4	4	5	.48308	.51807	.58509
4	4	6	.43516	.46650	.52740
4	4	7	.40047	.42889	.48459
4	4	8	.37403	.40009	.45145
4	4	9	.35312	.37723	.42494
4	4	10	.33609	.35857	.40318
4	5	5	.41518	.44546	.50461
4	5	6	.37028	.39699	.44969
4	5	7	.33815	.36211	.40966
4	5	8	.31388	.33566	.37905
4	5	9	.29481	.31482	.35480
4	5	10	.27938	.29792	.33506
4	6	6	.32786	.35118	.39755
4	6	7	.29774	.31850	.35997
4	6	8	.27512	.29387	.33147
4	6	9	.25743	.27457	.30903
4	6	10	.24316	.25899	.29085
5	5	5	.35158	.37745	.42853
5	5	6	.31050	.33307	.37790
5	5	7	.28142	.30148	.34153
5	5	8	.25963	.27775	.31401
5	5	9	.24262	.25918	.29239
5	5	10	.22893	.24421	.27490
5	6	6	.27240	.29190	.33086
5	6	7	.24562	.26284	.29738
5	6	8	.22567	.24113	.27222
5	6	9	.21016	.22422	.25256
5	6	10	.19771	.21064	.23673
6	6	6	.23660	.25348	.28727
6	6	7	.21217	.22701	.25679
6	6	8	.19407	.20734	.23403
6	6	9	.18006	.19209	.21634
6	6	10	.16887	.17990	.20214

Denote the cdf of l_{ih} by $G(t; i, h)$ and the pdf by $g(t; i, h)$. The distribution function for u_k is $G(t; 1, k-1)$ and, by Theorem 1, can be computed recursively from

$$G(t; i, k-1) = \int_{1/p_i}^{tq_i/p_k} G(t/z; i+1, k-1) \times g(z; i, i)\, dz \quad (3.2)$$

if $p_k/(p_i q_i) \le t \le p_k/q_i$,

$$G(t; i, k-1) = \int_{1/p_i}^{1} G(t/z; i+1, k-1) \times g(z; i, i)\, dz$$

if $p_k/q_i < t \le 1/p_i$, and

$$G(t; i, k-1) = 1 - \int_{t}^{1} [1 - G(t/z; i+1, k-1)] \times g(z; i, i)\, dz$$

if $1/p_i < t \le 1$.

Table 1 gives percentiles of u_3 when $m_1 \le m_2 \le 6$. The table was constructed by integrating (3.2) using the IMSL (1987) double precision subroutine DQDAPG on a VAX 8550. The functions $G(\cdot; i, i)$ and $g(\cdot; i, i)$ were obtained by the method in appendix 2 of Boik (1989a). For $p_i \le 3$, expressions for $G(\cdot; i, i)$ and $g(\cdot; i, i)$ are given, explicitly, in Davis (1972).

3.2 Nonnull Distributions

To motivate u_k as a statistic for testing $H_o : \lambda = 0$, we compare the nonnull distributions of u_k and the LRT statistic, w_k. The distribution of w_k is obtained by writing the residual sum of squares after fitting $\boldsymbol{\beta}$ and $\boldsymbol{\Delta}\lambda$ as a quadratic form in $\mathbf{e}$: $\text{SSE}(w_k) = \mathbf{e}'(\mathbf{I} - \hat{\mathbf{M}})\mathbf{e}$. The structure of $\hat{\mathbf{M}}$ follows from the normal Equations (2.4c):

$$\hat{\mathbf{M}} = \sum_{i=1}^{k} (\hat{\boldsymbol{\Delta}}_i \hat{\boldsymbol{\Delta}}_i' - \hat{\boldsymbol{\Delta}}\hat{\boldsymbol{\Delta}}') + \hat{\boldsymbol{\Delta}}\hat{\boldsymbol{\Delta}}'.$$

A minor modification of lemma 1 in Boik (1989b, p. 83) can be made to show that

$$|\hat{\boldsymbol{\delta}}_i \hat{\boldsymbol{\delta}}_i' - \boldsymbol{\delta}_i \boldsymbol{\delta}_i'| = O_p(\lambda^{-1})$$

for $i = 1, \ldots, k$. Suitably rearranging the terms in $(\mathbf{I} - \hat{\mathbf{M}})$ reveals that $(\mathbf{I} - \hat{\mathbf{M}})\mathbf{e} = (\mathbf{I} - \mathbf{M})\mathbf{e} + O_p(\lambda^{-1})$. The distribution of w_k, up to $O_p(\lambda^{-1})$, follows from the idempotency of $\mathbf{M}$. The distribution of u_k is obtained similarly. The results are summarized in Theorem 2.

Theorem 2. In Model (1.1) with (2.3),

$$[w_k/(1 - w_k)][(v - f)/f] \sim F(f, v - f, \lambda^2/\sigma^2) + O_p(\lambda^{-1}),$$

$$\text{SSE}(w_k)/\sigma^2 \sim \chi^2(v - f, 0) + O_p(\lambda^{-1}),$$

$$[u_k/(1 - u_k)][(v - f)/f] \sim F(f, v - f, \lambda^2/\sigma^2) + O_p(\lambda^{-1}),$$

and

$$\text{SSE}(u_k)/\sigma^2 \sim \chi^2(v - f, 0) + O_p(\lambda^{-1})$$

for w_k of (2.5), u_k of (3.1a), $\text{SSE}(x) = \mathbf{e}'\mathbf{e}(1 - x)$,

$$v = \prod_{i=1}^{k} (m_i - 1),$$

and

$$f = 1 - 2k + \sum_{i=1}^{k} m_i.$$

Theorem 2 reveals that the nonnull distributions of u_k and the LRT statistic are identical to within $O_p(\lambda^{-1})$. Consequently, the test based on u_k can be recommended as an approximation to the LRT. The noncentrality parameter in Theorem 2 is the best we could hope for, because it is the noncentrality parameter of the LRT statistic obtained assuming that $\boldsymbol{\Delta}$ is known. Theorem 2 also says that the nuisance parameters, $\boldsymbol{\Delta}$, play, at most, a minor role in the distributions of u_k and w_k. In fact, it can be shown that the distributions of the LRT and u_k statistics depend on σ^2 and $\boldsymbol{\Delta}\lambda$ solely through the noncentrality parameter λ^2/σ^2.

In a small simulation study with $k = 3$, the estimated power (at $\alpha = .05$ and $\alpha = .01$) agreed with the noncentral F computation to within one or two percentage points whenever power exceeded .10. This occurs, for example, when $|\lambda/\sigma| \ge 3$ and $m_i = 2$ for all i. For $m_1 = m_2 = 4$ and $m_3 = 10$, $|\lambda/\sigma| \ge 4$ is necessary.

3.3 Estimating σ^2

A variety of estimators of σ^2 in nonadditive two-way classifications have been proposed (Carter and Srivastava 1980; Hegemann and Johnson 1976; Johnson and Graybill 1972b; Mandel 1971; Marasinghe and Johnson 1982). In particular, Marasinghe (1985) and Schott (1986) showed that for large $|\lambda/\sigma|$, $\text{SSE}(w_2)/\sigma^2$ is approximately distributed as $\chi^2[(m_1 - 2)(m_2 - 2), 0]$. Marasinghe suggested $\hat{\sigma}^2 = \text{SSE}(w_2)/[(m_1 - 2)(m_2 - 2)]$ as an estimator of σ^2 if $H_o : \lambda = 0$ is rejected.

Theorem 2 suggests a k-way generalization of Marasinghe's estimator:

$$\hat{\sigma}^2 = \text{SSE}(w_k)/(v - f) \quad \text{or} \quad \tilde{\sigma}^2 = \text{SSE}(u_k)/(v - f) \quad (3.3)$$

for $\text{SSE}(\cdot)$, v, and f of Theorem 2 depending on whether the LRT or the u_k test is employed. The estimators in (3.3) are identical to Marasinghe's estimator when $k = 2$.

In the limiting case, $|\lambda| \to \infty$, the estimators in (3.3) are unbiased. For finite λ, the expected values of $\tilde{\sigma}^2/\sigma^2$ and $\hat{\sigma}^2/\sigma^2$ will be somewhat smaller than unity, with the bias of $\tilde{\sigma}^2$ being smaller than that of $\hat{\sigma}^2$. When $\lambda = 0$, percent bias is a function of the null expectations of $\{l_i\}$ in Theorem 1. For selected $\{m_i\}$, $\{E(l_i)\}$ can be obtained from the tables in Boik (1985). For the $\{m_i\}$ in Table 1 when $\lambda = 0$, percent bias of $\tilde{\sigma}^2$ ranges from -42% at (3, 3, 3) to -7% at (6, 6, 10). In a small simulation study of the three-way estimator, bias did not exceed ±5% provided that $|\lambda/\sigma| \geq 4$.

4. ILLUSTRATION

To illustrate the new methods, we will use the data collected by Batchelder, Rogers, and Walker (1966). The data also appear in John (1971, p. 85). Tobacco was grown in four soils, each treated with one of three mulches, over a period of three years. The response is yield in 100 pounds per acre. A comparison of the three mulches (no mulch, straw mulch, and straw mulch plus straw incorporated into the soil) was of primary interest. Batchelder et al. considered soil to be a blocking factor and analyzed the data according to a randomized block design. That is, they assumed that all interactions involving soil are 0. Table 2 gives the sums of squares required for such an analysis.

The computed value of u_3 for the tobacco data is .9237. The p value, from (3.2), is $\Pr(u_3 \geq .9237 \mid \lambda = 0) = .0040$ and thus, for $\alpha = .01$, $H_o : \lambda = 0$ is rejected. The computed value of the LRT statistic is $w_3 = .9238$. The exact p value for the LRT is unknown but obviously small. Thus both statistics suggest a nonadditive model. For comparison, the score test in Section 2.3 does not find evidence of three-way nonadditivity ($F = .44$, df = 1, 11). Apparently, three-way interaction is not simply related to the main effects.

From Theorem 2 and Equation (3.3), $\tilde{\sigma}^2 = .74$ with $v - f = 7$ df. Estimable functions of $\boldsymbol{\beta}$ can be tested using $\tilde{\sigma}^2$ as the denominator in the F statistics. A comparison of $\tilde{\sigma}^2$ to the mean squared residual in Table 2 (MSR = 5.62) suggests that tests performed assuming an additive model are strongly biased in favor of their respective null hypotheses.

Some experimenters might take rejection of $H_o : \lambda = 0$ as evidence that a transformation is needed. There is, however, no guarantee that a transformation to additivity can be found. In the Box–Cox (1964) power family, a power of .69 maximizes the likelihood function of the additive model. The 95% confidence interval for the power extends from .14 to 1.34, so the evidence that a transformation is needed is not very strong. Nevertheless, a square root transformation is suggested by the maximization. On square root transformed data, $u_3 = .81$ (p value = .07). Although not in the confidence interval, the log transformation does even better: $u_3 = .62$ (p value = .49). Thus, in this data set, transformable nonadditivity was detected by the proposed test. In data sets having nontransformable nonadditivity, inferences can be made on the original scale using $\tilde{\sigma}^2$.

Table 2. Analysis of Variance on Tobacco Yield

Source	*df*	*SS*	*MS*
Year	2	330.76	165.38
Mulch	2	167.70	83.85
Soil	3	57.74	19.25
Y*M	4	86.76	21.69
Y*S	6	21.26	3.54
M*S	6	28.03	4.67
Y*M*S	12	67.44	5.62

[*Received April 1988. Revised February 1989.*]

REFERENCES

Batchelder, A. R., Rogers, M. J., and Walker, J. P. (1966), "Effects of Subsoil Management Practices on Growth of Flue-Cured Tobacco," *Agronomy Journal*, 58, 345–347.

Boik, R. J. (1985), "A New Approximation to the Distribution Function of the Studentized Maximum Root," *Communications in Statistics—Simulation and Computation*, 14, 759–767.

——— (1989a), "Inference on Covariance Matrices Under Rank Restrictions," Technical Report 2-7-89, Montana State University, Dept. of Mathematical Sciences.

——— (1989b), "Reduced-Rank Models for Interaction in Unequally-Replicated Two-Way Classifications," *Journal of Multivariate Analysis*, 28, 69–87.

Box, G. E. P., and Cox, D. R. (1964), "An Analysis of Transformations" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 26, 211–246.

Carroll, J. D., and Chang, J. J. (1970), "Analysis of Individual Differences in Multidimensional Scaling via an *N*-Way Generalization of 'Eckart–Young' Decomposition," *Psychometrika*, 35, 283–319.

Carter, E. M., and Srivastava, M. S. (1980), "Asymptotic Distribution of the Latent Roots of the Noncentral Wishart Distribution and the Power of the Likelihood Ratio Test for Nonadditivity," *The Canadian Journal of Statistics*, 8, 119–134.

Carvalho, J. (1978), "Data Analysis of Multi-way Tables," *Dissertation Abstracts International*, 38, 3270B.

Davis, A. W. (1972), "On the Ratios of the Individual Latent Roots to the Trace of a Wishart Matrix," *Journal of Multivariate Analysis*, 2, 440–443.

Farmer, C. M. (1987), "A Study of Nonadditivity in Factorial Experiments With No Replication," *Dissertation Abstracts International*, 47, 4575B.

Harter, H. L., and Lum, M. D. (1962), "An Interpretation and Extension of Tukey's One Degree of Freedom for Non-Additivity," Technical Report 62-313, Aeronautical Research Laboratory, Wright-Patterson Air Force Base, Ohio.

Hegemann, V., and Johnson, D. E. (1976), "On Analyzing Two-Way AOV With Interaction," *Technometrics*, 18, 273–281.

International Mathematical and Statistical Libraries, Inc. (IMSL) (1987), *User's Manual: Math/Library*, Houston: Author.

James, A. T. (1954), "Normal Multivariate Analysis and the Orthogonal Group," *The Annals of Mathematical Statistics*, 25, 40–75.

John, P. W. M. (1971), *Statistical Design and Analysis of Experiments*, New York: Macmillan.

Johnson, D. E., and Graybill, F. A. (1972a), "An Analysis of a Two-Way Model With Interaction and No Replication," *Journal of the American Statistical Association*, 67, 862–868.

——— (1972b), "Estimation of σ^2 in a Two-Way Classification Model With Interaction," *Journal of the American Statistical Association*, 67, 388–394.

Kapteyn, A., Neudecker, H., and Wansbeek, T. (1986), "An Approach to *n*-Mode Components Analysis," *Psychometrika*, 51, 269–275.

Kettenring, J. R. (1983), "Components of Interaction in Analysis of Variance Models With No Replications," in *Contributions to Statistics: Essays in Honour of Norman L. Johnson*, ed. P. K. Sen, Amsterdam: North-Holland, pp. 283–297.

Krishnaiah, P. R., and Yochmowitz, M. G. (1980), "Inference on the Structure of Interaction in Two-Way Classification Model," in *Handbook of Statistics* (Vol. 1), ed. P. R. Krishnaiah, Amsterdam: North-Holland, pp. 973–994.

Krzanowski, W. J. (1979), "Some Exact Percentage Points of a Statistic Useful in the Analysis of Variance and Principal Components," *Technometrics*, 21, 261–263.

Mandel, J. (1971), "A New Analysis of Variance Model for Non-Additive Data," *Technometrics*, 13, 1–18.

Marasinghe, M. G. (1985), "Asymptotic Tests and Monte Carlo Studies Associated With the Multiplicative Interaction Model," *Communications in Statistics—Theory and Methods*, 14, 2219–2231.

Marasinghe, M. G., and Johnson, D. E. (1982), "Estimation of σ^2 in the Multiplicative Interaction Model," *Communications in Statistics—Theory and Methods*, 11, 315–324.

Milliken, G. A., and Graybill, F. A. (1970), "Extensions of the General Linear Hypothesis Model," *Journal of the American Statistical Association*, 65, 797–807.

Milliken, G. A., and Johnson, D. E. (1989), *Analysis of Messy Data, Volume 2: Nonreplicated Experiments*, New York: Van Nostrand.

SAS Institute Inc. (1985), *SAS/IML User's Guide*, Cary, NC: Author.

Schott, J. R. (1986), "A Note on the Critical Values Used in Stepwise Tests for Multiplicative Components of Interaction," *Communications in Statistics—Theory and Methods*, 15, 1561–1570.

Schuurmann, F. J., Krishnaiah, P. R., and Chattopadhyay, A. K. (1973), "On the Distribution of the Ratios of the Extreme Roots to the Trace of the Wishart Matrix," *Journal of Multivariate Analysis*, 3, 445–453.

Ten Berge, J. M. F., De Leeuw, J., and Kroonenberg, P. M. (1987), "Some Additional Results on Principal Components Analysis of Three-Mode Data by Means of Alternating Least Squares Algorithms," *Psychometrika*, 52, 183–191.

Tucker, L. R. (1966), "Some Mathematical Notes on Three-Mode Factor Analysis," *Psychometrika*, 31, 279–311.

Tukey, J. W. (1949), "One Degree of Freedom for Non-Additivity," *Biometrics*, 5, 232–242.