Computational Statistics & Data Analysis 16 (1993) 47–61
North-Holland

A three-degree of freedom test of additivity in three-way classifications

Mervyn G. Marasinghe
Iowa State University, Ames, IA 50011, USA

Robert J. Boik
Montana State University, Bozeman, MT 59717, USA

Received December 1990
Revised February 1992

Abstract: This article proposes a new interaction model for nonreplicated three-way classifications. A simulation study is used to show that a three-degree of freedom score test based on the new model compares favorably with existing one-degree of freedom score and likelihood ratio tests of additivity. The tests are illustrated through an analysis of a data set where it is shown how the new model may reveal a specific structure of three-factor interaction. This structure may be exploited to suggest possible explanations for the nonadditivity.

Keywords: Interaction; Multiplicative models; Nonadditivity; Two-way classification

1. Introduction

Several statistical procedures designed to detect nonadditivity in nonreplicated two-way classifications are discussed in the literature. While Tukey's (1949) single degree of freedom test is the most popular, the procedures of Mandel (1961, 1971) and Johnson and Graybill (1972) are also widely used. These and related methods are reviewed by Krishnaiah and Yochmowitz (1980) and Milliken and Johnson (1989).

Consider the model

$$y_{ij} = \mu + \alpha_i + \beta_j + \theta_{ij} + \epsilon_{ij}; \quad i = 1, \ldots, a; \; j = 1, \ldots, b; \tag{1}$$

where α_i, β_j, and θ_{ij} are the effects of two fixed factors and their interaction, respectively; and ϵ_{ij} are assumed to be i.i.d. $N(0, \sigma^2)$. The fixed effects are subject to the usual sum-to-zero restrictions. The previously mentioned procedures model the interaction term, θ_{ij}, as a function of other parameters.

Correspondence to: Dr. Mervyn Marasinghe, Dept. of Statistics, Iowa State University, Snedecor Hall, Ames, IA 50011-1210, USA.

Restricting the interaction to have a specific functional form enables one to partition the observed interaction into two components; namely the fit of the functional form and the complementary lack of fit. The lack of fit component is used to estimate σ^2, thereby making possible a test that the fit component has expectation zero. The resulting tests are naturally powerful when interaction exhibits the specific functional form assumed in the derivation. The tests are less powerful but still useful when interaction does not exhibit the specific functional form.

Scheffé (1959) showed that Tukey's test can be derived as a test of H_0: $\lambda = 0$ when the interaction is modeled as a product of the main effect parameters; i.e., $\theta_{ij} = \lambda\alpha_i\beta_j$. To motivate this model, Scheffé showed that if θ_{ij} is assumed to be a second-degree polynomial of α_i and β_j, then, as a consequence of the sum-to-zero restrictions, θ_{ij} is necessarily of the form $\lambda\alpha_i\beta_j$.

Milliken and Graybill (1970) proposed a general test that can be used whenever interaction is a known function of main effect parameters. The Milliken–Graybill test is exact because the test statistic has a central F distribution under additivity. In addition, St. Laurent (1990) showed that the Milliken–Graybill test is a score test. Accordingly, it is asymptotically equivalent to the corresponding likelihood ratio test. The tests proposed by Mandel (1961) and Tukey (1949, 1962) are special cases of the Milliken–Graybill test.

Mandel (1971) and Johnson and Graybill (1972) model two-way interaction as a product of two free parameters; i.e., $\theta_{ij} = \lambda\nu_i\xi_j$. This model assumes that the $a \times b$ matrix of interaction parameters has rank one. Johnson and Graybill (1972) showed that the likelihood ratio statistic for testing H_0: $\lambda = 0$ in this model is

$$U_1 = \hat{\lambda}^2/\text{tr}(\boldsymbol{Z}'\boldsymbol{Z}), \tag{2}$$

where $\boldsymbol{Z}$ is the $a \times b$ matrix of residuals from the additive model and $\hat{\lambda}^2$ is the maximum characteristic root of $\boldsymbol{Z}'\boldsymbol{Z}$.

Invariant tests that do not require specifying a functional form for $\{\theta_{ij}\}$ have been proposed by Boik (1990b, 1992) and Tusell (1990). These tests are sensitive to a wide class of alternatives but are less powerful than those that do assume a specific structure, provided that interaction does not deviate too far from the assumed structure.

The above methods have been very useful in studying interaction in two-way classifications. Their extensions to multiply-classified data, however, are not well-known and are not entirely straightforward. In this paper, we compare several methods for detecting interaction in three-way classifications including a new three-degree of freedom test.

We begin with the general model for nonadditive data

$$y_{ijk} = \mu + \alpha_i + \beta_j + \tau_k + \omega_{ij} + \nu_{ik} + \rho_{jk} + \theta_{ijk} + \epsilon_{ijk}, \tag{3}$$

for $i = 1, \ldots, a$; $j = 1, \ldots, b$; and $k = 1, \ldots, c$; and where it is assumed that all parameters except the error term are fixed and subject to the usual sum-to-zero restrictions. Subscripts i, j, and k represent levels of Factors A, B and C, respectively; and ϵ_{ijk} are assumed to be i.i.d. N(0, σ^2). Model (3) is saturated so that without imposing additional restrictions, σ^2 cannot be estimated. The conventional approach assumes that $\theta_{ijk}=0$. In the remainder of this paper, model (3) in which $\theta_{ijk}=0$ will be called the additive model (whether or not two-factor interactions exist). If, in fact, three-way interaction is nonexistent, then an unbiased estimator of σ^2 exists and main effects and two-factor interactions can be tested. If one of the factors is a blocking variable then the associated two-factor interactions may also be assumed to be zero. In Section 2, a one-degree of freedom score test and a likelihood ratio test based on a Mandel–Johnson–Graybill type model for θ_{ijk} are reviewed. In Section 3, we present a three-degree of freedom score test for three-way interaction. In Section 4 we compare the performance of these procedures using simulation and in Section 5, the procedures are illustrated with an analysis of a data set.

2. Current procedures for testing additivity in three-way classifications

As discussed in Section 1, three approaches have been employed for deriving tests of nonadditivity in two-way classifications. The first two model the interaction as a product of unknown parameters (main effect or free) while the third does not assume a specific structure. In this section, we review extensions of the first two approaches to the three-way case. The third approach has not been extended to multiply-classified data.

2.1. Harter and Lum's (1962) test for three-way interaction

Harter and Lum (1962) proposed that three-factor interaction in model (3) be expressed as a product of the main effect parameters. The Harter and Lum structure can be parameterized as $\theta_{ijk}=0$ if $(\boldsymbol{\alpha}'\boldsymbol{\alpha})(\boldsymbol{\beta}'\boldsymbol{\beta})(\boldsymbol{\tau}'\boldsymbol{\tau})=0$ and

$$\theta_{ijk}=\frac{\lambda\alpha_i\beta_j\tau_k}{[(\boldsymbol{\alpha}'\boldsymbol{\alpha})(\boldsymbol{\beta}'\boldsymbol{\beta})(\boldsymbol{\tau}'\boldsymbol{\tau})]^{1/2}}, \tag{4}$$

otherwise; where $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\tau}$ are a-, b-, and c-vectors of main effect parameters, respectively. The advantage of the parameterization in (4) over that in Harter and Lum (they omit division by the norm of the main effects) is that in (4) the magnitude of the interaction is indexed solely by λ rather than jointly by λ and the norm of the main effects. A single degree of freedom score test of H_0: $\lambda=0$ is readily obtained. Assume that the data are arranged as an abc-vector, $\boldsymbol{y}=\{y_{ijk}\}$; the elements of which are ordered such that index i changes slowest and index k changes fastest. That is,

$$\boldsymbol{y}=(y_{111}, y_{112},\ldots,y_{11c}, y_{121},\ldots,y_{12c},\ldots,y_{ab1},\ldots,y_{abc})'.$$

Following Milliken and Graybill (1970), a sum of squares for nonadditivity is

$$\mathrm{SS}_1 = \frac{\left[\left(\hat{\boldsymbol{\alpha}} \otimes \hat{\boldsymbol{\beta}} \otimes \hat{\boldsymbol{\tau}}\right)' y\right]^2}{(\hat{\boldsymbol{\alpha}}'\hat{\boldsymbol{\alpha}})\left(\hat{\boldsymbol{\beta}}'\hat{\boldsymbol{\beta}}\right)(\hat{\boldsymbol{\tau}}'\hat{\boldsymbol{\tau}})}, \tag{5}$$

where $\otimes$ is the Kronecker product operator; and $\hat{\boldsymbol{\alpha}}$, $\hat{\boldsymbol{\beta}}$, and $\hat{\boldsymbol{\tau}}$ are the usual least squares estimators of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\tau}$. The test statistic is

$$F_1 = \mathrm{SS}_1/\mathrm{MSE}_1, \tag{6}$$

where $\mathrm{MSE}_1 = (\mathrm{SS}_R - \mathrm{SS}_1)/(pqr-1)$; $p=a-1$; $q=b-1$; $r=c-1$; $\mathrm{SS}_R = z'z$; z is the abc-vector of residuals from model (3) with $\theta_{ijk}=0$ (i.e., the additive model); and $z_{ijk} = y_{ijk} - \bar{y}_{ij\cdot} - \bar{y}_{i\cdot k} - \bar{y}_{\cdot jk} + \bar{y}_{i\cdot\cdot} + \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdot\cdot k} - \bar{y}_{\cdot\cdot\cdot}$. Under H_0, F_1 is distributed as $F_{1,pqr-1}$.

2.2. *A likelihood ratio test for three-way interaction*

A flexible alternative to Harter and Lum's model is obtained by generalizing the Mandel–Johnson–Graybill model to three factors. The generalized model expresses θ_{ijk} as

$$\theta_{ijk} = \lambda\gamma_i\xi_j\delta_k, \tag{7}$$

where the vectors $\boldsymbol{\gamma}$, $\boldsymbol{\xi}$, and $\boldsymbol{\delta}$ each sum to zero and have unit norm. The alternative in (7) is a special case of Tucker's (1966) three-mode principal component model and Carroll and Chang's (1970) three-way singular value decomposition.

It is readily shown that the maximum likelihood estimators (MLEs) of λ^2, $\boldsymbol{\gamma}$, $\boldsymbol{\xi}$, $\boldsymbol{\delta}$ and σ^2 are given by the solutions to

$$\hat{\lambda}^2 = \max_{\boldsymbol{\gamma},\boldsymbol{\xi},\boldsymbol{\delta}} \left[(\boldsymbol{\gamma} \otimes \boldsymbol{\xi} \otimes \boldsymbol{\delta})' z\right]^2 \quad \text{and} \quad \hat{\sigma}^2 = (z'z - \hat{\lambda}^2)/abc, \tag{8}$$

where z is defined in (6). The MLEs of the remaining parameters are identical to those of the additive model. The normal equations corresponding to (8) are

$$\left(\hat{\boldsymbol{\gamma}} \otimes \hat{\boldsymbol{\xi}} \otimes \boldsymbol{I}_c\right)' z = \hat{\lambda}\hat{\boldsymbol{\delta}}, \qquad \left(\hat{\boldsymbol{\gamma}} \otimes \boldsymbol{I}_b \otimes \hat{\boldsymbol{\delta}}\right)' z = \hat{\lambda}\hat{\boldsymbol{\xi}}, \quad \text{and} \quad \left(\boldsymbol{I}_a \otimes \hat{\boldsymbol{\xi}} \otimes \hat{\boldsymbol{\delta}}\right)' z = \hat{\lambda}\hat{\boldsymbol{\gamma}}, \tag{9}$$

subject to the constraints on $\boldsymbol{\gamma}$, $\boldsymbol{\xi}$, and $\boldsymbol{\delta}$ stated earlier. The equations in (9) can be solved using an alternating least squares algorithm (Kroonenberg and de Leeuw, 1980). Under H_0: $\lambda=0$, the MLE of σ^2 is $(z'z)/abc$. Thus, the likelihood ratio test can be given as reject H_0 for large values of

$$U_3 = \hat{\lambda}^2/(z'z). \tag{10}$$

Boik (1990a) computed accurate percentage points for U_3 using a Jacobi polynomial expansion. Boik and Marasinghe (1989) proposed an approximation to the above test for which the required test statistic is easier to compute than U_3 and has a known null distribution.

Using Theorem 2 of Boik and Marasinghe (1989), under the alternative in (7) and for large $|\lambda/\sigma|$, the distribution of $(z'z-\hat{\lambda}^2)/\sigma^2$ can be approximated by a χ_g^2 distribution where $g=pqr-p-q-r+2$ for p, q, and r of (6). Thus if $\lambda=0$ is rejected under model (7), an estimate of experimental error variance is given by

$$\hat{\sigma}^2=(z'z-\hat{\lambda}^2)/g. \tag{11}$$

3. A new test for three-way interaction

In the absence of prior knowledge about the form of three-way interaction, we assume that θ_{ijk} can be modeled as a function of the main effects and two-factor interactions. Initially, assume (in the spirit of Scheffé) that θ_{ijk} can be approximated by a second-degree polynomial function of the parameters α_i, β_j, τ_k, ω_{ij}, ν_{ik}, and ρ_{jk}. Simplification, using the sum-to-zero restrictions, gives

$$\begin{aligned}\theta_{ijk}&=\phi_1\alpha_i\rho_{jk}+\phi_2\beta_j\nu_{ik}+\phi_3\tau_k\omega_{ij}+\phi_4\left(\rho_{jk}\nu_{ik}-c^{-1}\sum_{m=1}^{c}\rho_{jm}\nu_{im}\right)\\&\quad+\phi_5\left(\rho_{jk}\omega_{ij}-b^{-1}\sum_{m=1}^{b}\rho_{mk}\omega_{im}\right)+\phi_6\left(\nu_{ik}\omega_{ij}-a^{-1}\sum_{m=1}^{a}\nu_{mk}\omega_{mj}\right).\end{aligned} \tag{12}$$

Alternatively, one might consider three-way interaction as a manifestation of two-way interaction which differs over the levels of the third factor. For example, denote the BC interaction at the ith level of A by $\{\rho_{jk(i)}\}$. Averaging over the levels of Factor A gives the ordinary two-factor interaction: $\rho_{jk}=a^{-1}\sum_{i=1}^{a}\rho_{jk(i)}$. If the data are not additive, then $\{\rho_{jk(i)}\}\neq\{\rho_{jk}\}$ for some i. Thus, three-way interaction can be modeled by making the deviations between two-way interactions and their average over levels of a third factor depend on the levels of the third factor. If $\{\rho_{jk(i)}\}$ differs from $\{\rho_{jk}\}$ in magnitude only (not direction) and the deviations $\rho_{jk(i)}-\rho_{jk}$ are made to depend on i through the multiplicative parameter α_i, we obtain $\rho_{jk(i)}-\rho_{jk}=\phi_1\alpha_i\rho_{jk}$. Consideration of all two-way interactions results in three possible models for three-way interaction. Because it is not known, a priori, which of the three models best represents existing nonadditivity, we combine the three models giving the augmented model

$$\theta_{ijk}=\phi_1\alpha_i\rho_{jk}+\phi_2\beta_j\nu_{ik}+\phi_3\tau_k\omega_{ij}. \tag{13}$$

Model (13) can also be obtained from (12) by dropping the second-order terms involving the two-factor interaction parameters. Onukogu and Ama (1989) discuss an alternative model of this type. If desired, model (13) can be generalized by replacing α_i, β_j, and/or τ_k in (13) by free parameters, thus obtaining three-way generalizations of Mandel's (1961) bundle of lines and Tukey's (1962) vacuum cleaner models.

Using the general theory of Milliken and Graybill (1970), a three degree-of-freedom score test for testing H_0: $\phi_1 = \phi_2 = \phi_3 = 0$ in (13) can be derived. In practice, one can fit three covariates $\hat{h}_{1ijk}$, $\hat{h}_{2ijk}$, and $\hat{h}_{3ijk}$ in model (3) with $\theta_{ijk} = 0$, where $\hat{h}_{1ijk} = \hat{\alpha}_i \hat{\rho}_{jk}$, $\hat{h}_{2ijk} = \hat{\beta}_j \hat{\nu}_{ik}$, $\hat{h}_{3ijk} = \hat{\tau}_k \hat{\omega}_{ij}$ and $\hat{\alpha}_i$, $\hat{\beta}_j$ etc., are the usual least squares estimators of the corresponding parameters under the additive model. Alternatively, one can compute the sum of squares directly:

$$\mathrm{SS}_3 = \boldsymbol{y}'\hat{\boldsymbol{H}}(\hat{\boldsymbol{H}}'\hat{\boldsymbol{H}})^{-1}\hat{\boldsymbol{H}}'\boldsymbol{y}, \tag{14}$$

where $\hat{\boldsymbol{H}} = (\hat{\boldsymbol{h}}_1\ \hat{\boldsymbol{h}}_2\ \hat{\boldsymbol{h}}_3)$ is an $abc \times 3$ matrix; and $\hat{\boldsymbol{h}}_m = \{\hat{h}_{mijk}\}$, $m = 1$, 2, and 3 are abc-vectors. The test statistic is

$$F_3 = \mathrm{SS}_3/(3\mathrm{MSE}_3), \tag{15}$$

where $\mathrm{MSE}_3 = (\mathrm{SS}_R - \mathrm{SS}_3)/(pqr - 3)$ for p, q and r of (6). The Appendix contains a SAS program for computing F_3. By conditioning on $\hat{\boldsymbol{H}}$, it is readily shown that the null distribution of F_3 is $F_{3,pqr-3}$. Partial regression sums of squares can be used to test each product term in (13) using single degree-of-freedom F-tests. This will be illustrated in Section 5.

4. Comparison of power of three tests of additivity

In this section Harter and Lum's (1962) test (HL), the proposed three-degree of freedom test (3DF), and the likelihood ratio test (LR) are compared by means of simulation. Simulation studies were performed to compare the powers of the HL, LR, and 3DF tests under each of two alternative structures: the HL structure in (4) and the LR structure in (7). For the HL structure in (4), the three tests are compared under conditions which are optimal for the HL test. For the LR structure in (7), the three tests are compared under more general conditions. Before reporting the simulation results, some theoretical issues concerning the power of the three tests are reviewed.

4.1. Theoretical issues

To develop a general framework for the power study, model (3) can be expressed in matrix form as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\psi} + \boldsymbol{\theta} + \boldsymbol{\epsilon}, \tag{16}$$

where $\boldsymbol{X}$ is the $abc \times t$ design matrix coding for main effects and two-factor interactions, $\mathrm{rank}(\boldsymbol{X}) = u$, $u = abc - (a-1)(b-1)(c-1)$, $\boldsymbol{\psi}$ is the t-vector of main effects and two-factor interaction parameters and $\boldsymbol{\theta}$ is the abc-vector of three-factor interaction parameters.

For the additive model, that is when $\boldsymbol{\theta} = \boldsymbol{0}$, the usual least squares estimator of $\boldsymbol{\psi}$ is $\hat{\boldsymbol{\psi}} = (\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y}$. It can be shown that conditional on $\boldsymbol{X}\hat{\boldsymbol{\psi}}$, statistics (6) and (15) have doubly noncentral F distributions of the form

$$F_{s,abc-u-s,\hat{d},d-\hat{d}}, \tag{17}$$

where $\hat{d} = \boldsymbol{\theta}'\hat{\boldsymbol{G}}(\hat{\boldsymbol{G}}'\hat{\boldsymbol{G}})^{-}\hat{\boldsymbol{G}}'\boldsymbol{\theta}/\sigma^2$, $d = \boldsymbol{\theta}'\boldsymbol{\theta}/\sigma^2$, $s = \text{rank}(\hat{\boldsymbol{G}})$, and $\hat{\boldsymbol{G}}$ takes different forms depending on the test statistic under consideration. This result follows from arguments similar to those used by Ghosh and Sharma (1963) and Hegemann and Johnson (1976) when computing the power of Tukey's test. For the HL test, $\hat{\boldsymbol{G}}$ is the *abc*-vector $\hat{\boldsymbol{\alpha}} \otimes \hat{\boldsymbol{\beta}} \otimes \hat{\boldsymbol{\tau}}$, and for the 3DF test $\hat{\boldsymbol{G}}$ is $\hat{\boldsymbol{H}}$ in (14).

The power of the HL and 3DF tests can be computed for any specified structure, $\boldsymbol{\theta}$, by evaluating the expectation

$$\text{Power} = \text{E}\left\{\Pr\left[F_i \geq F^{1-\alpha}_{i,abc-u-i} \mid \hat{d}\right]\right\}, \tag{18}$$

where $F^{1-\alpha}_{i,abc-u-i}$ is the $100(1-\alpha)$ percentile of the F distribution with i and $abc-u-i$ degrees of freedom; F_i for $i = 1$ and 3 are given in (6) and (15); and the expectation is taken with respect to the distribution of $\hat{d}$. To estimate power, $\hat{d}$ is computed for randomly generated data sets having a fixed value for $\boldsymbol{\psi}$. The doubly noncentral F distribution in (17) is evaluated for each value of $\hat{d}$ and then averaged over data sets to obtain an accurate estimate of power.

Conditional on $\boldsymbol{X}\hat{\boldsymbol{\psi}}$, the power of the score test (HL or 3DF) is a monotonic function of the numerator noncentrality parameter, $\hat{d}$. Accordingly, for the unconditional power of the score test (defined in Eq. (18)) to be high, the distribution of $\hat{d}$ must assign high probability to large values of $\hat{d}$ and low probability to small values of $\hat{d}$. The exact distribution of $\hat{d}$ is quite messy, but some simple approximations can be obtained. In particular, in the HL test against (4), $\hat{d}$ can be written as

$$\hat{d} = d(1-\eta) \tag{19}$$

where

$$\eta = \max\left\{\text{O}_p\left[\left(\frac{bc\boldsymbol{\alpha}'\boldsymbol{\alpha}}{\sigma^2}\right)^{-1/2}\right], \text{O}_p\left[\left(\frac{ac\boldsymbol{\beta}'\boldsymbol{\beta}}{\sigma^2}\right)^{-1/2}\right], \text{O}_p\left[\left(\frac{ab\boldsymbol{\tau}'\boldsymbol{\tau}}{\sigma^2}\right)^{-1/2}\right]\right\},$$

and where the random term, η, has support on the interval [0, 1]. Equation (19) reveals that for fixed d, the HL test has greatest power against (4) when $bc\boldsymbol{\alpha}'\boldsymbol{\alpha}$, $ac\boldsymbol{\beta}'\boldsymbol{\beta}$, and $ab\boldsymbol{\tau}'\boldsymbol{\tau}$ are each large relative to σ^2. Accordingly, for fixed total main effect magnitude,

$$bc\boldsymbol{\alpha}'\boldsymbol{\alpha} = ac\boldsymbol{\beta}'\boldsymbol{\beta} = ab\boldsymbol{\tau}'\boldsymbol{\tau} \tag{20}$$

is optimal for the HL test.

For fixed d and main effect directions [e.g., $\boldsymbol{\alpha}(\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1/2}$], the power of the 3DF test against (4) is maximized by choosing two-way interactions such that with high probability, $d - \hat{d}$ is small. One way to do this is to choose a two-way interaction, such that $\text{E}(\hat{\boldsymbol{h}}_m)$ in (14) is proportional to $\boldsymbol{\theta}$ in (4), for $m = 1, 2,$ *or* 3. Power may be even higher (but not lower) if all three two-way interactions are selected in this manner. Accordingly, the optimal two-factor interaction structures for the 3DF test are

$$\boldsymbol{\rho} = w_1 \frac{\boldsymbol{\beta}\boldsymbol{\tau}'}{[(\boldsymbol{\beta}'\boldsymbol{\beta})(\boldsymbol{\tau}'\boldsymbol{\tau})]^{1/2}}, \qquad \boldsymbol{\nu} = w_2 \frac{\boldsymbol{\alpha}\boldsymbol{\tau}'}{[(\boldsymbol{\alpha}'\boldsymbol{\alpha})(\boldsymbol{\tau}'\boldsymbol{\tau})]^{1/2}}, \quad \text{and}$$

$$\boldsymbol{\omega} = w_3 \frac{\boldsymbol{\alpha}\boldsymbol{\beta}'}{[(\boldsymbol{\alpha}'\boldsymbol{\alpha})(\boldsymbol{\beta}'\boldsymbol{\beta})]^{1/2}}. \tag{21}$$

It can be shown that if (21) is satisfied, then $\hat{d}$ in the 3DF test against (4) can be written as

$$\hat{d} = d(1-\eta), \tag{22}$$

where

$$\eta \le \min\left\{ \mathrm{O}_p\left[\left(\frac{bc\boldsymbol{\alpha}'\boldsymbol{\alpha}}{\sigma^2}\right)^{-1/2}\right] + \mathrm{O}_p\left[\left(\frac{aw_1^2}{\sigma^2}\right)^{-1}\right], \right.$$
$$\mathrm{O}_p\left[\left(\frac{ac\boldsymbol{\beta}'\boldsymbol{\beta}}{\sigma^2}\right)^{-1/2}\right] + \mathrm{O}_p\left[\left(\frac{bw_2^2}{\sigma^2}\right)^{-1}\right],$$
$$\left. \mathrm{O}_p\left[\left(\frac{ab\boldsymbol{\tau}'\boldsymbol{\tau}}{\sigma^2}\right)^{-1/2}\right] + \mathrm{O}_p\left[\left(\frac{cw_3^2}{\sigma^2}\right)^{-1}\right] \right\},$$

and where the random term, η, has support on the interval [0, 1].

Equation (22) reveals that if d is fixed and (21) is satisfied, then the 3DF test has greatest power against (4) when both elements of at least one pair, $[bc\boldsymbol{\alpha}'\boldsymbol{\alpha}, aw_1^2]$, $[ac\boldsymbol{\beta}'\boldsymbol{\beta}, bw_2^2]$, or $[ab\boldsymbol{\tau}'\boldsymbol{\tau}, cw_3^2]$ are large relative to σ^2. Thus, for sufficiently large d, the 3DF test can have acceptable power as long as the magnitudes of at least one main effect and the complementary two-way interaction are large. When (21) holds, the power of the 3DF test depends only on w_1, w_2, w_3, d, and the magnitude of the main effects.

It can be shown that against (4), the power of the LR test depends on σ^2 and $\boldsymbol{\theta}$ solely through d in (17). Theorem 2 in Boik and Marasinghe (1989) can be used to approximate the power of the LR test. Nevertheless, for comparability to the HL and 3DF results, simulation will be used.

4.2. Simulation results

In each simulation study, $5 \times 5 \times 5$ tables of data were randomly generated according to model (3) using different sets of values for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\tau}$, $\boldsymbol{\omega}$, $\boldsymbol{\nu}$, and $\boldsymbol{\rho}$ [as well as $\boldsymbol{\gamma}$, $\boldsymbol{\xi}$, and $\boldsymbol{\delta}$ in the case of (7)]. Main effect parameters were chosen to satisfy (20) because this is an optimal choice for the HL test against (4). Two-factor interaction parameters were chosen to satisfy (21) with $w_1 = w_2 = w_3$. While the structures in (21) are optimal for the 3DF test against (4), equating two-way interaction magnitudes and imposing (20) are each sub-optimal. For fixed total main effect magnitude (say $\boldsymbol{\alpha}'\boldsymbol{\alpha} + \boldsymbol{\beta}'\boldsymbol{\beta} + \boldsymbol{\tau}'\boldsymbol{\tau} = \kappa_1\sigma^2$) and fixed total two-way interaction magnitude (say $w_1^2 + w_2^2 + w_3^2 = \kappa_2\sigma^2$), 3DF power is maximized by setting w_i^2 to $\kappa_2\sigma^2$ for $i = 1, 2,$ *or* 3, and letting the squared norm of the complementary main effect approach $\kappa_1\sigma^2$. The HL test would have minimal power under these conditions.

For each of three values of d in (17), namely 4, 16, and 32, 1200 simulation trials were performed. Table 1 summarizes the results when the interaction has the HL structure in (4). Table 1 shows that the power of each test against (4) increases as d increases. For fixed d, the power of HL increases as $(\boldsymbol{\alpha}'\boldsymbol{\alpha})/\sigma^2$ increases and dominates LR if $(\boldsymbol{\alpha}'\boldsymbol{\alpha})/\sigma^2$ is large enough. The dependence of the power of 3DF on the main effect magnitude is less than that of HL, which shows only moderate increase with increasing $(\boldsymbol{\alpha}'\boldsymbol{\alpha})/\sigma^2$. On the other hand, the power of 3DF increases with w_1^2/σ^2 for fixed d and exceeds that of HL if w_1^2/σ^2 is large enough and $(\boldsymbol{\alpha}'\boldsymbol{\alpha})/\sigma^2$ is not too large.

Estimated power of HL and 3DF against the LR structure in (7) is displayed in Figure 1. In each plot, main effects and two-factor interactions are fixed and $\boldsymbol{\gamma}$, $\boldsymbol{\xi}$, and $\boldsymbol{\delta}$ are varied such that the squared distance

$$L^2 = \min_{\pi} \left\| \frac{\boldsymbol{\alpha} \otimes \boldsymbol{\beta} \otimes \boldsymbol{\tau}}{\sqrt{(\boldsymbol{\alpha}'\boldsymbol{\alpha})(\boldsymbol{\beta}'\boldsymbol{\beta})(\boldsymbol{\tau}'\boldsymbol{\tau})}} - \pi(\boldsymbol{\gamma} \otimes \boldsymbol{\xi} \otimes \boldsymbol{\delta}) \right\|^2$$

$$= 1 - \frac{[(\boldsymbol{\gamma}'\boldsymbol{\alpha})(\boldsymbol{\xi}'\boldsymbol{\beta})(\boldsymbol{\delta}'\boldsymbol{\tau})]^2}{(\boldsymbol{\alpha}'\boldsymbol{\alpha})(\boldsymbol{\beta}'\boldsymbol{\beta})(\boldsymbol{\tau}'\boldsymbol{\tau})} \qquad (23)$$

assumes different values in the interval [0, 1]. Under the LR alternative (7), the nonnull distribution of the LR statistic depends on σ^2 and $\boldsymbol{\theta}$ solely through $d = \lambda^2/\sigma^2$.

In Figure 1, the powers of HL, 3DF and LR against (7) for values of $d = 4$, 16 and 32 are plotted as a function of L defined in (23). All three tests increase in power with increasing d, but to different degrees. The powers of HL and 3DF increase as L decreases; i.e., as $\boldsymbol{\gamma} \otimes \boldsymbol{\xi} \otimes \boldsymbol{\delta}$ becomes proportionally closer to $\boldsymbol{\alpha} \otimes \boldsymbol{\beta} \otimes \boldsymbol{\tau}$. The power of LR, of course, is constant over L. Thus, in the top three panels, the power of HL exceeds that of 3DF for all (L, d) and both HL and 3DF beat LR unless L and d are large. In the bottom three panels, 3DF beats HL for all (L, d) and 3DF beats LR unless L and d are large.

In summary, each of the 3 tests has the ability to detect three-way nonadditivity under different conditions. For fixed d, the power of the HL test under (4) depends totally on the size of the main effects; power may be low even if only one of the main effects is small (see Eq. (19)). The 3DF test has the advantage of a reduced dependence on the main effects (see Eq. (22)). The LR test does not depend at all on the main effects. Thus the 3DF test strikes a balance between HL and LR. It is less dependent on main effects than is HL but at a cost of 2 df. It is less general than LR and this saves $a + b + c - 2$ df.

Table 1
Estimated power of three tests for three-way interaction against $\theta_{ijk} = \lambda\alpha_i\beta_j\tau_k$ ($\alpha = 0.05$)

$(\boldsymbol{\alpha}'\boldsymbol{\alpha})/\sigma^2$	w_1^2/σ^2	LR	HL	3DF
d = 4.0				
0.125	1.0	0.0649	0.09	0.09
	2.0			0.11
	4.0			0.12
0.25	1.0		0.12	0.10
	2.0			0.12
	4.0			0.13
0.5	1.0		0.17	0.10
	2.0			0.12
	4.0			0.14
1.0	1.0		0.21	0.11
	2.0			0.13
	4.0			0.15
d = 16.0				
0.125	1.0	0.2934	0.14	0.20
	2.0			0.28
	4.0			0.36
0.25	1.0		0.29	0.24
	2.0			0.33
	4.0			0.42
0.5	1.0		0.48	0.28
	2.0			0.37
	4.0			0.46
1.0	1.0		0.63	0.30
	2.0			0.40
	4.0			0.48
d = 32.0				
0.125	1.0	0.7508	0.22	0.36
	2.0			0.50
	4.0			0.62
0.25	1.0		0.45	0.44
	2.0			0.59
	4.0			0.70
0.5	1.0		0.72	0.51
	2.0			0.66
	4.0			0.70
1.0	1.0		0.88	0.55
	2.0			0.70
	4.0			0.80

5. An example

In this section we consider an example to illustrate the three tests. Table 2 presents data first published in Xhonga (1971) and later used by Brown (1975). The experiment involved the measurement of hardness of gold fillings made

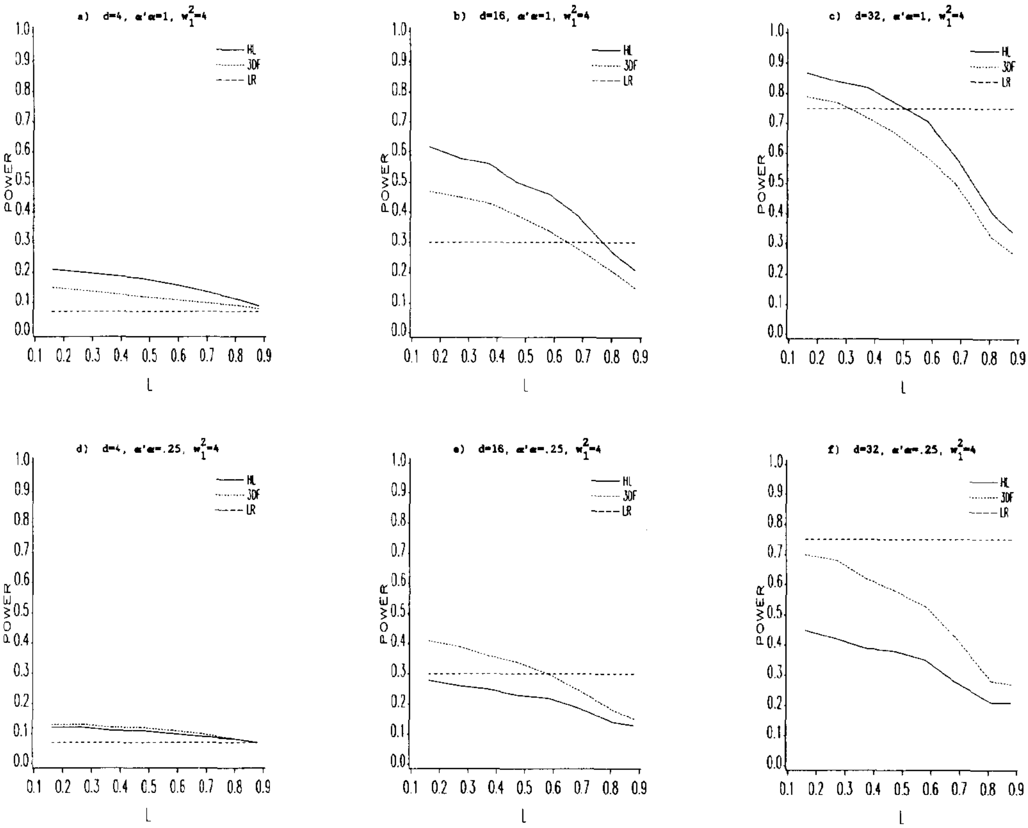

Fig. 1. Power of HL, 3DF, and LR tests against $\theta_{ijk}=\lambda\gamma_i\xi_j\delta_k$ with $\alpha=0.05$, $\sigma^2=1$, $L^2=1-[(\gamma'\alpha)(\xi'\beta)(\delta'\tau)]^2/(\alpha'\alpha)(\beta'\beta)(\tau'\tau)$, and $d=\lambda^2/\sigma^2$.

using 8 types of gold (Type) and 3 methods of condensation (Cond) by 5 dentists (Dent). Table 3 gives an analysis of variance on the data, scaled by dividing each value by 100.

The computed value of SS_1 in (5) is 2.19, giving a value of 2.25 for F_1 which is not significant at 5%. Thus, HL fails to detect three-way interaction. However, SS_2 in (14) is computed to be 10.03 which gives a value of 3.87 for F_3, significant at 5%. The computed value of $\hat{\lambda}^2$ in (8) is 27.92 giving a value of 0.5 for U_3 in (10). The 5% critical value is approximately 0.4464, and hence, H_0: $\lambda=0$ is rejected. Thus, 3DF and LR tests find evidence of three-way nonadditivity. An estimate of σ^2 using (11) is $\tilde{\sigma}^2=0.62$ with 45 df.

At this stage one could attempt a transformation to achieve additivity. In the Box–Cox power family, a value of 2.1 maximizes the likelihood function with a 95% confidence interval of (1.6, 2.7). For the transformed data, using a power of 2, $F_3=0.18$ while $U_3=0.4617$, the second still significant at 5%. It may be that the nonadditivity in this data is not transformable. Brown (1975) suggested that the Cond × Dent interaction can be attributed to the 10 smallest observations in the data set. A referee has pointed out that the attempt to remove non-additivity by a Box–Cox transformation is clearly unreasonable because the transfor-

Table 2
The hardness of gold fillings: 8 types of gold (Type)×3 methods of condensation (Cond)×5 dentists (Dent)

Dentist	Method of condensation	Gold							
		1	2	3	4	5	6	7	8
1	1	792	824	813	792	792	907	792	835
	2	772	772	782	698	665	115	835	870
	3	782	803	752	620	835	847	560	585
2	1	803	803	715	803	813	858	907	882
	2	752	772	772	782	743	933	792	824
	3	715	707	835	715	673	698	734	681
3	1	715	724	743	627	752	858	762	724
	2	792	715	813	743	613	824	847	782
	3	762	606	743	681	743	715	824	681
4	1	673	946	792	743	762	894	792	649
	2	657	743	690	882	772	813	870	858
	3	690	245	493	707	289	715	813	312
5	1	634	715	707	698	715	772	1048	870
	2	694	724	803	665	752	824	933	835
	3	724	627	421	483	405	536	405	312

mation simply corrects for the negative skew induced by these small values. To examine whether the three-factor interaction might also be an artifact of these observations, they were deleted and the HL and 3DF statistics recomputed. Neither test is significant, perhaps supporting the above conjecture.

A more plausible explanation of the significance of U_3 is obtained by examining the structure of three-factor interaction. An analysis of the three components of the 3DF sum of squares is summarized in Table 4. Table 4 suggests that differences in the Type × Dent interaction among levels of Cond are proportional to the Cond main effects. An examination of the marginal means for Cond reveals that the Cond main effect is primarily due to a difference between Cond3 and (Cond1 + Cond2)/2. The interaction between this contrast and Type × Dent accounts for most of the three-factor interaction. The results of this analysis are summarized in Table 5. For further interpreta-

Table 3
Analysis of variance for gold filling data

Source	DF	SS	MS
Dent	4	21.76	5.44
Cond	2	59.76	29.88
Type	7	22.03	3.15
Dent × Cond	8	26.34	3.29
Dent × Type	28	20.88	0.75
Dent × Type	14	20.98	1.50
Dent × Cond × Type	56	55.83	1.00

Table 4
Partitioning of three-degree of freedom sum of squares for non-additivity for the gold data (note that $\hat{h}_1$, $\hat{h}_2$ and $\hat{h}_3$ were defined near (14))

Source	DF	SS	MS	F	p-value
Dent × Cond × Type	56	55.83	1.00		
Regression ($\hat{h}_1$, $\hat{h}_2$, $\hat{h}_3$)	3	10.03			
$\hat{h}_1$	1	2.95	2.95	3.41	0.07
$\hat{h}_2$ after $\hat{h}_1$	1	5.46	5.46	6.32	0.02
$\hat{h}_3$ after $\hat{h}_1$, $\hat{h}_2$	1	1.62	1.62	1.87	0.18
$\hat{h}_2$	1	5.64	5.64	6.42	0.01
$\hat{h}_3$ after $\hat{h}_2$	1	1.12	1.12	1.30	0.25
$\hat{h}_1$ after $\hat{h}_2$, $\hat{h}_3$	1	3.27	3.27	3.78	0.06
$\hat{h}_3$	1	2.15	2.15	2.49	0.12
$\hat{h}_1$ after $\hat{h}_3$	1	3.61	3.61	4.20	0.04
$\hat{h}_2$ after $\hat{h}_1$, $\hat{h}_3$	1	4.27	4.27	4.94	0.03
Residual	53	45.8	0.86		

tion of the structure uncovered by this post-hoc analysis, one must appeal to subject matter considerations.

Appendix. SAS commands to perform 3DF test

```
Options LS=80;
Data;
*;
*    Data are in file Gold.Dat;
*;
     Infile Gold;
     Input Cond Dent Type Hard;
```

Table 5
Sums of squares partitioned by (Cond1 + Cond2)/2 vs. Cond3

Comparison	DF	SS	MS	F	p-value
(Cond1 + Cond2)/2 vs. Cond3	1	59.76	59.76	131.75	0.0001
Interaction with Type	7	18.20	2.60	5.73	0.0003
Interaction with Dent	4	25.73	6.43	14.18	0.0001
Interaction with Type × Dent	28	43.12	1.54	3.40	0.0009
Residual	28	12.70	0.54		

```
*;
*    Fit Models to Construct 3 Covariates;
*    X1, X2, and X3;
*;
Proc GLM noprint;
     Class Dent Cond Type;
     Model Hard"="Cond Dent|Type;
     Output P=X1;
Proc GLM noprint;
     Class Dent Cond Type;
     Model Hard"="Cond|Type Dent;
     Output P=X2;
Proc GLM noprint;
     Class Dent Cond Type;
     Model Hard"="Cond|Dent Type;
     Output P=X3;
Data;
     set work.data4;
     X1=X1**2/2;
     X2=X2**2/2;
     X3=X3**2/2;
*;
*    Fit and Test 3DF for Nonadditivity;
*;
Proc GLM;
     Class Dent Cond Type;
     Model Hard"="Cond|Dent|Type@2 X1 X2 X3;
     Estimate '"X1: BC x A ' X1 1;
     Estimate '"X2: AC x B ' X2 1;
     Estimate '"X3: AB x C ' X3 1;
     Contrast '"3DF ' X1 1, X2 1, X3 1;
```

References

Boik, R.J., A likelihood ratio test for three-mode singular values: Upper percentiles and an application to three-way ANOVA. *Computational Statistics and Data Analysis* **10** (1990a) 1–9.

Boik, R.J., Inference on covariance matrices under rank restrictions. *Journal of Multivariate Analysis* **33** (1990b) 230–246.

Boik, R.J., Testing additivity in two-way classifications with no replications: The locally best invariant test. *Journal of Applied Statistics* (1992) In press.

Boik, R.J., and M.G. Marasinghe, Analysis of nonadditive multi-way classifications. *Journal of American Statistical Association* **84** (1989) 1059–1064.

Brown, M.B., Exploring interaction effects in the ANOVA. *Applied Statistics* **24** (3) (1975) 288–298.

Carroll, J.D. and J.J. Chang, Analysis of individual differences in multidimensional scaling via an *N*-way generalization of 'Eckart–Young' decomposition. *Psychometrika* **35** (1970) 283–319.

Farmer, C.M., A Study of Nonadditivity in Factorial Experiments with No Replication. Unpublished Ph.D. Dissertation, Iowa State University, Department of Statistics, (1986).

Ghosh, M.N. and D. Sharma, Power of Tukey's test for nonadditivity. *Journal of the Royal Statistical Society* **25B** No. 1 (1963) 213–219.

Harter, H.L. and M.D. Lum, *An Interpretation and Extension of Tukey's One Degree of Freedom for Non-Additivity*. Aeronautical Research Laboratory Technical Report, ARL 62-313, (1962).

Hegemann V. and D.E. Johnson, The power of two tests for nonadditivity. *Journal of American Statistical Association* **71** (1976) 945–948.

Johnson, D.E. and F.A. Graybill, An analysis of a two-way model with interaction and no replication. *Journal of the American Statistical Association* **67** (1972) 862–868.

Krishnaiah, P.R. and M.G. Yochmowitz, Inference on the structure of interaction in two-way classification model. In: P.R. Krishnaiah (Ed.) *Handbook of Statistics (Vol. 1)* (North Holland, Amsterdam, 1980) 973–994.

Kroonenberg, P.M. and J. de Leeuw, Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika* **45** (1980) 69–97.

Mandel, J., Non-additivity in two-way analysis of variance. *Journal of the American Statistical Association* **56** (1961) 878–888.

Mandel, J., A new analysis of variance model for non-additive data. *Technometrics* **13** (1971) 1–18.

Milliken, G.A. and F.A. Graybill, Extensions of the general linear hypothesis model. *Journal of the American Statistical Association* **65** (1970) 797–807.

Milliken, G.A. and D.E. Johnson, *Analysis of Messy Data, Vol. 2: Non-replicated Experiments* (Von Nostrand Reinhold, New York, 1989).

Onukogu, I.B. and N.O. Ama, Two models for interaction in a 3-way analysis of variance with one observation per cell. *Biometrical Journal* **31** (1989) 649–662.

Scheffé, H., *The analysis of Variance* (Wiley, New York, 1959).

St. Laurent, R.T., The equivalence of the Milliken–Graybill procedure and the score test. *The American Statistician* **44** (1990) 36–37.

Tusell, F., Testing for interaction in two-way ANOVA tables with no replication, *Computational Statistics and Data Analysis*. **10** (1990) 29–45.

Tucker, L.R., Some mathematical notes on three-mode factor analysis. *Psychometrika* **31** (1966) 279–311.

Tukey, J.W., One degree of freedom for nonadditivity. *Biometrics* **5** (1949) 232–242.

Tukey, J.W., The future of data analysis. *Annals of Mathematical Statistics* **33** (1962) 1–67.

Xhonga, F., Direct gold alloys-Part II. *Journal of the American Academy of Gold Foil Operators* **14** (1971) 5–15.