Computational Statistics & Data Analysis 10 (1990) 1–9
North-Holland

A likelihood ratio test for three-mode singular values: Upper percentiles and an application to three-way ANOVA

Robert J. Boik

Montana State University, Bozeman, MT 59717, USA

Received May 1989
Revised November 1989

Abstract: This paper considers the rank-1 three-mode model for an $n_2 n_3 \times n_1$ matrix, $\boldsymbol{Y}$. In vector form, the model is $\boldsymbol{y} = (\boldsymbol{v}_1 \otimes \boldsymbol{v}_2 \otimes \boldsymbol{v}_3)\lambda + \boldsymbol{\epsilon}$, where $\boldsymbol{y} = \mathrm{vec}(\boldsymbol{Y})$, $\boldsymbol{v}_j$ is an $n_j \times 1$ vector of parameters, $\boldsymbol{v}_j'\boldsymbol{v}_j = 1$ for $j = 1, 2, 3$, and $\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$. The likelihood ratio test of $H_0: \lambda = 0$ is given and, employing a Jacobi polynomial expansion, upper percentiles of the null distribution of the test statistic are computed. As an illustration, the results are applied to the problem of testing additivity in unreplicated three-way classifications.

Keywords: Multimode principal components, multimode singular value decomposition, multiplicative model, nonadditivity, reduced-rank model.

1. Introduction

The approximation of a matrix, $\boldsymbol{Y}: n_2 \times n_1$, by a lower rank matrix has a long history in statistics. The model underlying the reduced-rank approximation can be written as

$$\boldsymbol{Y} = \boldsymbol{M} + \boldsymbol{E}, \tag{1}$$

where $\boldsymbol{M}$ is restricted to have rank-r, $r < \min(n_1, n_2)$, and $\boldsymbol{E}$ is a matrix of random residuals having expectation $\boldsymbol{0}$ and dispersion $\mathrm{var}[\mathrm{vec}(\boldsymbol{E})] = \sigma^2 \boldsymbol{I}$. Eckart and Young [6] showed that the least squares estimator of $\boldsymbol{M}$ is what we now call the rank-r singular value decomposition (SVD) of $\boldsymbol{Y}$. That is, $\hat{\boldsymbol{M}} = \hat{\boldsymbol{V}}_2 \hat{\boldsymbol{\Lambda}} \hat{\boldsymbol{V}}_1'$ where $\hat{\boldsymbol{\Lambda}} = \mathrm{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_r)$; $\hat{\lambda}_i$ is the positive square root of the ith largest characteristic root of $\boldsymbol{Y}'\boldsymbol{Y}$; $\hat{\boldsymbol{V}}_1$ is an $n_1 \times r$ orthonormal matrix of rank-r ($\boldsymbol{V}_1'\boldsymbol{V}_1 = \boldsymbol{I}_r$) consisting of the corresponding characteristic vectors of $\boldsymbol{Y}'\boldsymbol{Y}$; and $\hat{\boldsymbol{V}}_2 = \boldsymbol{Y}\hat{\boldsymbol{V}}_1\hat{\boldsymbol{\Lambda}}^{-1}$.

An extension of SVD models to three-dimensional data was proposed by Tucker [17]. Law, Snyder, Hattie, and McDonald [14] give an overview of

research related to Tucker's three-mode model. Tucker's model was generalized to four dimensions by Lastovicka [13] and to an arbitrary number of dimensions by Kapteyn, Neudecker, and Wansbeek [11]. Following Magnus and Neudecker ([15], §17.11), the k-mode SVD model can be motivated by writing $\boldsymbol{M}$ in (1) as $\boldsymbol{M} = \boldsymbol{V}_2 \boldsymbol{\Lambda} \boldsymbol{V}_1'$ and expressing the two-mode model in vector form:

$$y = (\boldsymbol{V}_1 \otimes \boldsymbol{V}_2)\lambda + \epsilon, \tag{2}$$

where $y = \text{vec}(\boldsymbol{Y})$, $\lambda = \text{vec}(\boldsymbol{\Lambda})$, and $\epsilon = \text{vec}(\boldsymbol{E})$. The k-mode extension, suggested by (2), is

$$y = (\boldsymbol{V}_1 \otimes \cdots \otimes \boldsymbol{V}_k)\lambda + \epsilon, \tag{3}$$

where y is an n-vector; $n = \Pi_{j=1}^{k} n_j$; $\boldsymbol{V}_j$ is $n_j \times r_j$, has rank-r_j, and is orthonormal; λ is an r-vector; and $r = \Pi_{j=1}^{k} r_j$. Each element of y is associated with a unique k-tuple of indices; the jth index runs from 1 to n_j. The elements in y are ordered such that $n_1 \leq \cdots \leq n_k$; the first index changes slowest; and the kth index changes fastest etc.

Note that it is the fixed effects in model (3), rather than the random effects, which have a k-mode structure. The covariance matrix has a 1-mode structure: $\text{var}(\epsilon) = \sigma^2 \boldsymbol{I}_n$. Multimode models for random effects have also been proposed (Bentler, Poon, & Lee [1]). In these models, the covariance matrix has a k-mode structure while the fixed effects structure is 1-mode.

If the random errors in (3) follow a spherical normal distribution, $\epsilon \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$, then maximum likelihood estimators of the unknown parameters are

$$\hat{\sigma}^2 = \frac{y'y - \hat{\lambda}'\hat{\lambda}}{n}, \tag{4a}$$

where $\hat{\lambda}$ along with $\hat{\boldsymbol{V}}_j$ for $j = 1, \ldots, k$ are the minimizers of

$$\| y - (\boldsymbol{V}_1 \otimes \cdots \otimes \boldsymbol{V}_k)\lambda \|^2. \tag{4b}$$

The minimization in (4b) can be accomplished efficiently using the alternating least squares algorithm of Kapteyn, Neudecker, and Wansbeek [11]. Inference procedures for the parameters in model (3), however, are not as well developed. Boik and Marasinghe [2], in an ANOVA context, constructed an exact test of $H_0 : \lambda = 0$ for a special case of the k-mode SVD model:

$$y = (\boldsymbol{v}_1 \otimes \boldsymbol{v}_2 \otimes \boldsymbol{v}_3)\lambda + \epsilon, \tag{5}$$

where $\boldsymbol{v}_j$ is $n_j \times 1$ for $j = 1, 2, 3$. Their test, however, is not the likelihood ratio test (LRT). It employs a non-maximum likelihood based statistic which was proposed, primarily, to circumvent difficult distribution problems associated with maximum likelihood based statistics.

This paper is concerned with the LRT of $H_0 : \lambda = 0$ in model (5). The test is easily shown to be the following: reject H_0 for large values of

$$l = 1 - \hat{\phi}^{2/n} = \frac{\hat{\lambda}^2}{y'y}, \tag{6}$$

where $\hat{\phi}$ is the ratio of the likelihood function maximized under H_0 to the

likelihood function maximized under $H_a: \lambda \neq 0$ and $\hat{\lambda}^2$ is the maximum likelihood estimator of λ^2. The exact null distribution of l is, unfortunately, unknown. Farmer [7] used simulation to estimate the null moments of l and approximated the distribution of l by a beta distribution. Farmer's upper percentiles are useful as initial approximations, but are not especially accurate and cover a limited range of (n_1, n_2, n_3) triplets: $n_1 = 2 \leq n_2 \leq 5$ and $n_2 \leq n_3 \leq 8$. The primary contribution of this paper is to compute accurate upper percentiles of the null distribution of the LRT statistic for a larger set of (n_1, n_2, n_3) triplets. Computational details are described in §2. An application to nonreplicated three-way classifications is given in §3.

2. Computational details

2.1. A lower bound statistic

The Boik and Marasinghe non-LRT statistic for testing H_0 is of interest in the present setting because its distribution can be used as a crude approximation to that of the LRT statistic. This Section describes the crude approximation which, in section 2.2, is improved through the use of orthogonal polynomials.

The Boik and Marasinghe statistic is computed as follows. First, a one-step (non-maximum likelihood) estimate of $\boldsymbol{v}_1$, say $\tilde{\boldsymbol{v}}_1$, is computed. Second, the likelihood function, with $\tilde{\boldsymbol{v}}_1$ substituted for $\boldsymbol{v}_1$, is maximized with respect to the remaining unknown parameters. The test statistic is given by

$$u = 1 - \tilde{\phi}^{2/n} = \frac{\tilde{\lambda}_1^2 \tilde{\lambda}_2^2}{\boldsymbol{y}'\boldsymbol{y}}, \tag{7}$$

where $\tilde{\phi}$ is the ratio of the likelihood function maximized under H_0 to the likelihood function approximately maximized under H_a; $\tilde{\lambda}_1^2$ is the maximum characteristic root of $\boldsymbol{Y}'\boldsymbol{Y}$; $\boldsymbol{Y}$ is the $n_2 n_3 \times n_1$ matrix satisfying $\text{vec}(\boldsymbol{Y}) = \boldsymbol{y}$; $\tilde{\lambda}_2^2$ is the maximum characteristic root of $\boldsymbol{Y}_1'\boldsymbol{Y}_1$; $\boldsymbol{Y}_1$ is the $n_3 \times n_2$ matrix satisfying $\text{vec}(\boldsymbol{Y}_1) = \boldsymbol{Y}\tilde{\boldsymbol{v}}_1\tilde{\lambda}_1^{-1}$; and $\tilde{\boldsymbol{v}}_1$, is the unit-norm characteristic vector associated with the maximum root, $\tilde{\lambda}_1^2$. It is readily shown that $u \in [(n_1 n_2)^{-1}, 1]$ with probability 1. Theorem 1 in Boik and Marasinghe [2] gives the exact null distribution of u.

Denote the cumulative distribution functions (cdfs) of l and u by $F_l(x)$ and $F_u(x)$, respectively. By construction, u is stochastically smaller than l. That is, $F_u(x) \geq F_l(x)$ for all real x. Thus, $F_l(x)$ can be written as

$$F_l(x) = F_u(x) - \delta(x), \tag{8}$$

where $\delta(x) = F_u(x) - F_l(x) \geq 0$ for all real x.

A crude approximation to the cdf of l is provided by $F_u(x)$. The error in this approximation is $\delta(x)$ and is illustrated in Fig. 1 for the case $(n_1, n_2, n_3) = (2, 2, 3)$ and in Fig. 2 for the case $(n_1, n_2, n_3) = (5, 5, 9)$. Each figure summarizes the results of a small simulation study. Uniform random variates were generated using a 64 component compound random number generator of the type described

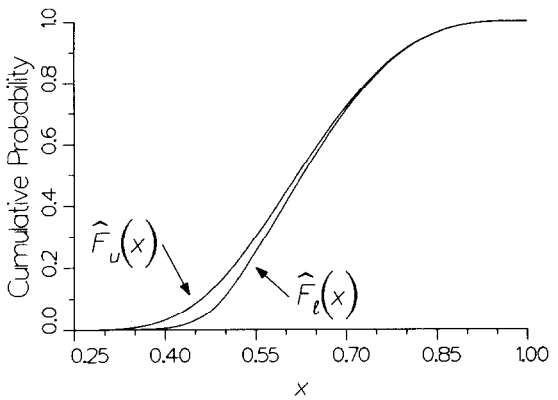

Fig. 1. Empirical cdfs based on a 50,000 trial simulation: $(n_1, n_2, n_3) = (2, 2, 3)$.

by Collings [5]. The compound generator is based on multipliers taken from the list given by Fishman and Moore [8]. The uniform variates were transformed to random normal variates by the Box and Muller [3] method. For each of 50,000 pseudorandom vectors, y, the statistics l and u were computed. Fig. 1 shows that, for $(n_1, n_2, n_3) = (2, 2, 3)$, the empirical cdfs differ mainly for x values corresponding to percentiles smaller than 50. The maximum difference, $\hat{F}_u(x) - \hat{F}_l(x)$, occurs at $x = 0.482$ and is only about 6.75%. From Fig. 2, it is apparent that the cdfs of u and l differ much more for the (5, 5, 9) case than for the (2, 2, 3) case. The empirical cdfs differ for all percentiles except those very near 0% or 100%. The maximum difference occurs at $x = 0.138$ and exceeds 37%.

2.2. *Jacobi polynomial expansion of* $\delta(x)$

A more accurate approximation to $F_l(x)$ can be obtained by adjusting $F_u(x)$ by an approximation to $\delta(x)$. By the Weierstrass approximation theorem, $\delta(x)$ can be expanded as $\delta(x) = p_N(x) + \epsilon_N(x)$ where $p_N(x)$ is a polynomial of degree N

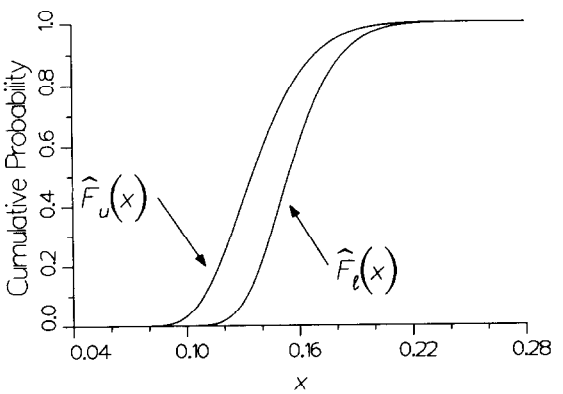

Fig. 2. Empirical cdfs based on a 50,000 trial simulation: $(n_1, n_2, n_3) = (5, 5, 9)$.

and, by choosing N to be sufficiently large, $\max_x |\epsilon_N(x)|$ can be made arbitrarily small. Of course, to obtain an accurate approximation with small N, the polynomials must be carefully chosen. Prior research (Farmer [7], Johnson & Graybill [10]) has shown that the cdfs of statistics such as l and u (i.e., statistics computed as maximum root ÷ trace) can be approximated reasonably well by beta distributions. The present application makes a milder assumption: the cdfs of l and u can each be accurately approximated by a finite weighted sum of beta cdfs. Under this assumption, $\delta(x)$ is also accurately approximated by a finite weighted sum of beta cdfs. Accordingly, $\delta(x)$ will be expanded in terms of Jacobi polynomials (Hochstrasser [9]) with weight function $w(x) = x^{\alpha-1}(1-x)^{\beta-1}$.

Define z by

$$z = x - \frac{(1-x)}{n_1 n_2 - 1}, \tag{9}$$

so that $z \in [0, 1]$ for $x \in [(n_1 n_2)^{-1}, 1]$. Write $\delta(x)$ as

$$\delta(x) = \rho f(z), \qquad \text{where } f(z) = \frac{1}{\rho}\left[F_u\left(z + \frac{1-z}{n_1 n_2}\right) - F_l\left(z + \frac{1-z}{n_1 n_2}\right)\right], \tag{10}$$

$\rho = n_1 n_2 E(l-u)/(n_1 n_2 - 1)$ and $E(\cdot)$ is the expectation operator. Define μ_i for $i = 0, \ldots, \infty$ by

$$\mu_i = \int_0^1 z^i f(z)\, \mathrm{d}z.$$

It can be shown that

$$\mu_i = \frac{(n_1 n_2)^i}{(n_1 n_2 - 1)^i E(l-u)} \sum_{j=0}^{i} \binom{i}{j} (-1)^j \frac{E(l^{i-j+1} - u^{i-j+1})}{(i-j+1)(n_1 n_2)^j}. \tag{11}$$

Note that $f(z)$ is a density function; it integrates to 1 and is nonnegative for $z \in [0, 1]$.

Because the domain of $f(z)$ is finite, the sequence of moments, $\mu_0, \ldots, \mu_\infty$, uniquely determines the density function (Rao [16], §2b.5). Hence, $f(z)$ can be written as an infinite degree Jacobi polynomial. Matching the first three moments, μ_0, μ_1, and μ_2, to those of a beta distribution gives the three-term Jacobi expansion:

$$f(z) \approx \frac{\Gamma(\gamma+\beta) z^{\gamma-1}(1-z)^{\beta-1}}{\Gamma(\gamma)\Gamma(\beta)}, \tag{12}$$

where $\gamma = \mu_1(\mu_1 - \mu_2)/(\mu_2 - \mu_1^2)$ and $\beta = (1-\mu_1)(\mu_1 - \mu_2)/(\mu_2 - \mu_1^2)$ for μ_i of (11). The complete expansion of $F_l(x)$, based on equations (8) to (12) is

$$F_l(x) = F_u(x) \tag{13}$$

$$-\rho \lim_{N\to\infty} \sum_{i=0}^{N} \omega_i \sum_{j=0}^{i} \tau_{ij} \left(x - \frac{1-x}{n_1 n_2 - 1} \right)^{\gamma+i-j-1} \left(\frac{(1-x)n_1 n_2}{n_1 n_2 - 1} \right)^{\beta-1}, \qquad \text{where}$$

$$\omega_i = \frac{(\gamma+\beta+2i-1)\Gamma(\gamma+i)}{i!\Gamma(\gamma+\beta+i-1)\Gamma(\beta+i)} \sum_{j=0}^{i} \tau_{ij}\mu_{i-j}, \qquad \text{and}$$

$$\tau_{ij} = \binom{i}{j}(-1)^j \frac{\Gamma(\gamma+\beta+2i-j-1)}{\Gamma(\gamma+i-j)}.$$

2.3. Computation of the upper percentiles

Computation of the upper percentiles requires two levels of approximation. First, the series in (13) must be truncated by fixing N. In the present application, the series was truncated at $N = 10$. Second, for fixed N, evaluation of (13) requires that $E(l^i - u^i)$ for $i = 1, \ldots, N+1$ be known. These expectations for $2 \le n_1 \le n_2 \le 5$, and $n_2 \le n_3 \le 9$ were estimated by simulation. For each (n_1, n_2, n_3) triplet, 50,000 pseudo-random normal vectors, y, were generated by the method described in 2.1. For each realization of y, l and u were computed. The expectations $E(l^i - u^i)$ for $i = 1, \ldots, 11$ were estimated by the means of the observed values. Employing these estimated expectations in (10) to (13) yields $\hat{F}_l(x)$.

Fig. 3 and 4 display the resulting estimates of $\delta(x)$ for the $(n_1, n_2, n_3) = (2, 2, 3)$ and $(n_1, n_2, n_3) = (5, 5, 9)$ cases. The differences between the empirical cdfs, from Fig. 1 and 2, are shown for comparison. The Jacobi expansions displayed in Fig. 3 and 4 are based on moments estimated independently from the empirical distribution functions displayed in Fig. 1 and 2. For both cases, the Jacobi expansion of $\delta(x)$ closely approximates the empirical cdf based estimate. In Fig. 3, the maximum difference between the two estimates is only about $\frac{1}{3}\%$. This differences occurs at $x = 0.518$ which corresponds, approximately, to the 16th percentile of l. Differences are much smaller at the upper percentiles, which are of much more interest. In Fig. 4, the maximum difference between the two estimates is only about 0.6%. This difference occurs at $x = 0.137$ which also

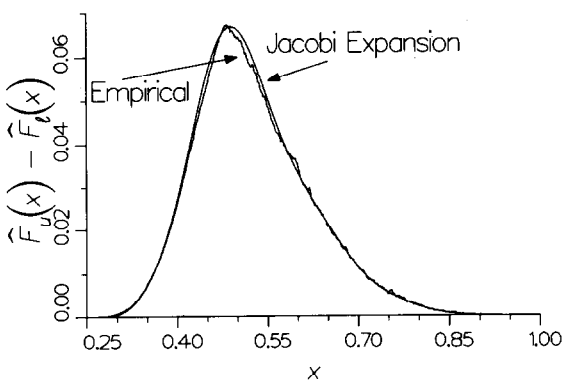

Fig. 3. Difference between cdfs: $(n_1, n_2, n_3) = (2, 2, 3)$.

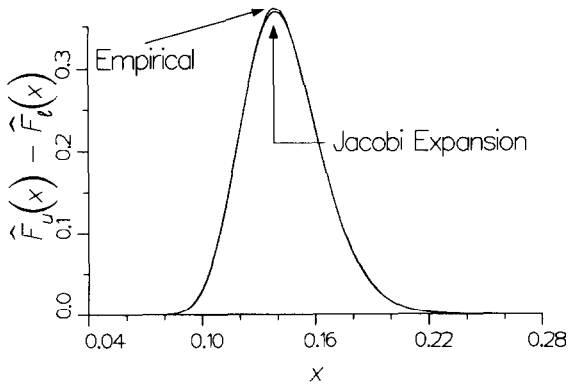

Fig. 4. Difference between cdfs: $(n_1, n_2, n_3) = (5, 5, 9)$.

corresponds, approximately, to the 16th percentile of l. As for the (2, 2, 3) case, differences are much smaller at the upper percentiles. Together, Fig. 3 and 4 suggest that upper percentiles based on the Jacobi expansion ought to be quite accurate.

Upper percentiles were computed by equating $\hat{F}_l(x)$ to $1-\alpha$ for selected α and solving for x. The results are given in Table 1.

3. An application

This section illustrates the LRT of $H_0: \lambda = 0$ for testing additivity in a non-replicated three-way classification model. The linear model for classifications having n_1, n_2, and n_3 levels can be written as $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\theta} + \boldsymbol{\epsilon}$ where $\boldsymbol{y}$ is an $(n_1 n_2 n_3)$-vector, $\boldsymbol{X}$ is the design matrix coding for main effects and two-factor interactions, and $\boldsymbol{\theta}$ is a vector of three-factor interactions, subject to the usual sum to zero constraints. To develop a test for additivity, Boik and Marasinghe [2] modeled $\boldsymbol{y} - \boldsymbol{X\beta}$ as

$$\boldsymbol{y} - \boldsymbol{X\beta} = (\boldsymbol{v}_1 \otimes \boldsymbol{v}_2 \otimes \boldsymbol{v}_3)\lambda + \boldsymbol{\epsilon}. \tag{14}$$

Model (14) is a generalization of the Johnson–Graybill [10] multiplicative model for interaction. A variant of model (14) is described and illustrated in Kettenring [12].

The maximum likelihood estimator of $\boldsymbol{X\beta}$ can be written as $\boldsymbol{X\hat{\beta}} = \boldsymbol{Hy}$ where

$$\boldsymbol{H} = \boldsymbol{I}_n - [(\boldsymbol{I}_1 - \boldsymbol{H}_1) \otimes (\boldsymbol{I}_2 - \boldsymbol{H}_2) \otimes (\boldsymbol{I}_3 - \boldsymbol{H}_3)],$$

$\boldsymbol{I}_i$ is an $n_i \times n_i$ identity, $\boldsymbol{H}_i = n_i^{-1}\boldsymbol{J}_i$, and $\boldsymbol{J}_i$ is an $n_i \times n_i$ matrix of ones. Making the substitution, $\boldsymbol{X\hat{\beta}}$ for $\boldsymbol{X\beta}$ in (14), the model to be fit is

$$\boldsymbol{e} = (\boldsymbol{v}_1 \otimes \boldsymbol{v}_2 \otimes \boldsymbol{v}_3)\lambda + \boldsymbol{\eta}, \tag{15}$$

where $\boldsymbol{e} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{y}$ is the vector of residuals from the additive fit and $\mathrm{Var}(\boldsymbol{e}) = \sigma^2[(\boldsymbol{I}_1 - \boldsymbol{H}_1) \otimes (\boldsymbol{I}_2 - \boldsymbol{H}_2) \otimes (\boldsymbol{I}_3 - \boldsymbol{H}_3)]$.

Table 1
Upper percentiles of the likelihood ratio statistic

n_1	n_2	n_3	$1-\alpha$			n_1	n_2	n_3	$1-\alpha$		
			0.90	0.95	0.99				0.90	0.95	0.99
2	2	2	0.8896	0.9237	0.9668	3	3	7	0.3851	0.4091	0.4578
2	2	3	0.7918	0.8336	0.8989	3	3	8	0.3642	0.3867	0.4315
2	2	4	0.7241	0.7661	0.8374	3	3	9	0.3471	0.3679	0.4101
2	2	5	0.6754	0.7157	0.7874	3	4	4	0.4261	0.4546	0.5105
2	2	6	0.6383	0.6765	0.7465	3	4	5	0.3822	0.4067	0.4570
2	2	7	0.6090	0.6454	0.7127	3	4	6	0.3500	0.3722	0.4169
2	2	8	0.5852	0.6196	0.6846	3	4	7	0.3258	0.3459	0.3868
2	2	9	0.5653	0.5982	0.6607	3	4	8	0.3070	0.3251	0.3627
2	3	3	0.6777	0.7200	0.7940	3	4	9	0.2911	0.3079	0.3429
2	3	4	0.6053	0.6445	0.7167	3	5	5	0.3405	0.3620	0.4054
2	3	5	0.5555	0.5912	0.6593	3	5	6	0.3107	0.3293	0.3677
2	3	6	0.5182	0.5513	0.6151	3	5	7	0.2878	0.3047	0.3399
2	3	7	0.4895	0.5201	0.5798	3	5	8	0.2695	0.2849	0.3170
2	3	8	0.4662	0.4947	0.5512	3	5	9	0.2551	0.2697	0.2996
2	3	9	0.4472	0.4741	0.5275	4	4	4	0.3648	0.3885	0.4360
2	4	4	0.5329	0.5677	0.6347	4	4	5	0.3245	0.3450	0.3864
2	4	5	0.4835	0.5149	0.5765	4	4	6	0.2954	0.3137	0.3506
2	4	6	0.4478	0.4763	0.5319	4	4	7	0.2737	0.2900	0.3234
2	4	7	0.4201	0.4464	0.4982	4	4	8	0.2563	0.2712	0.3020
2	4	8	0.3975	0.4218	0.4703	4	4	9	0.2426	0.2563	0.2839
2	4	9	0.3799	0.4025	0.4478	4	5	5	0.2870	0.3044	0.3400
2	5	5	0.4354	0.4633	0.5185	4	5	6	0.2601	0.2756	0.3073
2	5	6	0.4007	0.4258	0.4756	4	5	7	0.2398	0.2534	0.2820
2	5	7	0.3741	0.3966	0.4424	4	5	8	0.2241	0.2366	0.2625
2	5	8	0.3527	0.3738	0.4166	4	5	9	0.2110	0.2224	0.2466
2	5	9	0.3355	0.3551	0.3944	5	5	5	0.2520	0.2670	0.2975
3	3	3	0.5615	0.5996	0.6704	5	5	6	0.2277	0.2408	0.2674
3	3	4	0.4923	0.5252	0.5897	5	5	7	0.2093	0.2210	0.2447
3	3	5	0.4457	0.4748	0.5330	5	5	8	0.1948	0.2054	0.2270
3	3	6	0.4116	0.4378	0.4903	5	5	9	0.1832	0.1928	0.2125

To obtain parameter estimates for (15), first transform from $\boldsymbol{e}$ to $\boldsymbol{z} = (\boldsymbol{K}_1 \otimes \boldsymbol{K}_2 \otimes \boldsymbol{K}_3)'\boldsymbol{e}$ where $\boldsymbol{K}_i\boldsymbol{K}_i' = \boldsymbol{I}_i - \boldsymbol{H}_i$ and $\boldsymbol{K}_i$ is $n_i \times (n_i - 1)$. The distribution of $\boldsymbol{z}$ is nonsingular with $\text{var}(\boldsymbol{z}) = \sigma^2\boldsymbol{I}$ so that the model for $\boldsymbol{z}$ can be fit by ordinary least squares. Thus, the net effect of substituting $\boldsymbol{X}\hat{\boldsymbol{\beta}}$ for $\boldsymbol{X}\boldsymbol{\beta}$ is to reduce the dimensions from (n_1, n_2, n_3) to $(n_1 - 1, n_2 - 1, n_3 - 1)$.

As an example, Table 1 in Brown [4] gives the results of an experiment concerned with the hardness of dental fillings constructed from gold. Five dentists used each of three methods of condensation to construct fillings from each of eight types of gold. The experiment was conducted by Xhonga [18]. The observed value of Boik and Marasinghe's statistic, u, is 0.4896. The corresponding p-value is 0.0113 so that $H_0: \lambda = 0$ is rejected for $\alpha = 0.05$. The computed LRT statistic is $l = 0.5002$. The critical value, from Table 1, using $\alpha = 0.01$ and

$(n_1, n_2, n_3) = (2, 4, 7)$, is 0.4982 so that H_0 is rejected for $\alpha = 0.01$. The Jacobi expansion gives a p-value of 0.0093.

References

[1] P.M. Bentler, W.Y. Poon, and S.Y. Lee, Generalized multimode latent variable models: Implementation by standard programs, *Computational Statistics and Data Analysis* **7** (1988) 107–118.

[2] R.J. Boik and M.G. Marasinghe, Analysis of nonadditive multiway classifications, *Journal of the American Statistical Association* **84** (1989) 1059–1064.

[3] G.E.P. Box and M.E. Muller, A note on generation of normal deviates, *Annals of Mathematical Statistics* **29** (1958) 610–611.

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

[5] B.J. Collings, Compound random number generators, *Journal of the American Statistical Association* **82** (1987) 525–527.

[6] C. Eckart and G. Young, The approximation of one matrix by another of lower rank, *Psychometrika* **1** (1936) 211–218.

[7] C.M. Farmer, A study of nonadditivity in factorial experiments with no replication (Doctoral dissertation, Iowa State University, 1986) *Dissertation Abstracts International* **47** (1987) 4575B.

[8] G.S. Fishman and L.R. Moore, An exhaustive analysis of multiplicative congruental random number generators with modulus $2^{31} - 1$, *SIAM Journal on Scientific and Statistical Computing* **7** (1986) 24–45; Corrigenda **7** (1986) 1058.

[9] U.W. Hochstrasser, Orthogonal polynomials, in: M.A. Abramowitz and I.A. Stegun (Eds.), *Handbook of Mathematical Functions* (Dover, New York, 1965).

[10] D.E. Johnson 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.

[11] A. Kapteyn, H. Neudecker, and T. Wansbeek, An approach to n-mode components analysis, *Psychometrika* **51** (1986) 269–275.

[12] J.R. Kettenring, Components of interaction in analysis of variance models with no replication, in P.K. Sen (ed.) *Contributions to Statistics: Essays in Honour of Norman L. Johnson* (North-Holland, Amsterdam, 1983).

[13] J.L. Lastovicka, The extension of component analysis to four-mode matrices, *Psychometrika* **46** (1981) 47–57.

[14] H.G. Law, C.W. Hattie, and R.P. McDonald, *Research Methods for Multimode Data Analysis* (Praeger, New York, 1984).

[15] J.R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics* (Wiley, New York, 1988).

[16] C.R. Rao, *Linear Statistical Inference and its Applications, Second Edition* (Wiley, New York, 1973).

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

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