THREE-MODE PRINCIPAL COMPONENT ANALYSIS IN DESIGNED EXPERIMENTS

by
Kyoungah See

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY
in
Statistics

APPROVED:

Eric P. Smith

Klaus Hinkelmann

Marvin Lentner

Raymond H. Myers

George Terrell

March 31, 1993
Blacksburg, Virginia

THREE-MODE PRINCIPAL COMPONENT ANALYSIS IN DESIGNED EXPERIMENTS

by

Kyoungah S. See

Eric P. Smith, Chair

Statistics

(ABSTRACT)

This dissertation is concerned with the application of a method for decomposing multivariate data, three-mode principal component analysis, to a three-way table with one observation per cell. It is based on the class of multiplicative models for three-way tables ($s \times t \times u$) whose general form has expectation

$$E(y_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \sum_{p=1}^{s}\sum_{q=1}^{t}\sum_{r=1}^{u} c_{pqr}\, g_{ip}\, h_{jq}\, e_{kr}.$$

The application is related to tests for nonadditivity in the two-way analysis of variance with noreplication.

Three-factor interaction can be assessed for three-way cross classified tables with only one observation per treatment combination. This is done by partitioning the three-factor interaction sum of squares into a portion related to the interaction and a portion associated with random error. In particular, the estimated interaction matrix is decomposed by three-mode principal component analysis to separate significant interaction from random error. Three test procedures are presented for assessing interaction: randomization tests, Monte Carlo methods, and likelihood ratio tests. Examples illustrating the use of these approaches are presented.

In addition to the above testing approaches, a graphical procedure, joint plots, is investigated for diagnosing the type of model to fit to three-way arrays of data. The plot is a multiway analogue of the biplot graphical analysis for two-way matrices. Each observation is represented by a linear combination of inner products of markers which are obtained from three-mode principal component analysis. The relationship between various models and the geometrical configurations of the plots on Euclidean spaces of such markers allows one to diagnose the type of model which fits the data. An example is given to illustrate the simplicity of the technique and the usefulness of this graphical approach in diagnosing models.

This dissertation is dedicated to
my mother, Junghee Jo,
and to
the memory of my father, Byung Gil See.

Acknowledgements

First, I would like to express my most heartfelt thanks to my advisor, Dr. Eric Smith, for his superb guidance and support. He introduced me to three-mode principal component analysis and provided me with constant encouragement. He made my research career at Virginia Tech most enjoyable, being easy to get along with, and always enthusiastic about what I was doing. I am forever grateful for the experiences I take with me.

I would like to thank my committee members: Drs. Hinkelmann, Lentner, Myers, and Terrell. Especially, I am deeply grateful to Dr. Klaus Hinkelmann for his careful review of my dissertation and many suggestions of changes and constructive criticism. I also want to sincerely thank Dr. Marvin Lentner, for his willingness to serve on my committee and for the assistance he gave me in completing my program; Dr. Ray Myers, for his confidence in me and his valuable support and advice; and to Dr. George Terrell, for his thought-provoking questions which gave me more insight on my research problems.

My appreciation is extended to all my professors of the Department of Statistics at Virginia Tech for providing an excellent education, especially to Dr. Don Jensen, for his expertise and his valuable suggestions on the materials in Section 3.4 of this dissertation.

Last, but far from least, to my families in Seoul, Korea and in the United States, your love and support has meant so much to me.

Table of Contents

List of Tables

List of Illustrations

CHAPTER 1. INTRODUCTION

A standard approach, in testing data from two-factor replicated studies, is to test first for interaction, then assess main effects if there is no interaction and do follow-up analysis if there is interaction. However, this approach is difficult when there is no replication. When there is only one observation per cell, there is no way to directly test for interaction since there is no independent measure of error. The interaction mean square is made up of both the inherent variability and the true interaction effect if any. Attempts have been made to establish models to partition the interaction sums of squares into something related to interactions and something more associated with random error. One approach to do this was due to Tukey (1949) who split off a single degree of freedom associated with a test for nonadditivity.

An alternative test for nonadditivity may be obtained using principal component analysis, or rather singular value decomposition. The test applies the Eckart and Young theorem (1936) on the least squares property of the singular value decomposition of the matrix of the interaction terms estimated from the data. For two-way tables, this approach has been studied by many researchers (Gollob (1968), Mandel (1971), Johnson and Graybill (1972), Cortsen and Van Eijnsbergen (1972), Krishnaiah and Yochmowitz (1980)). These authors examined the following model (usually associated with Mandel, 1971)

$$y_{ij} = \mu + \alpha_i + \beta_j + \sum_{p=1}^{r} \theta_p u_{pi} v_{pj} + \epsilon_{ij},$$

which has been employed to analyze an unreplicated two-way table.

There are also several ways of informally implementing this singular value decomposition approach using various graphical plots such as, interaction plots (Snee (1982), Milliken and Johnson (1989)), probability plots (Kettenring, (1983)), and biplots (Gabriel (1971), Bradu and Gabriel (1978)).

A three-way classification model is a useful model in experimental investigations. To make a complete model, it is necessary to determine if all two and three factor interactions are present and also to get an estimate of the error variance σ^2. Since the data we discuss has exactly one observation per cell, the conventional linear model theory cannot be used to check for the three-factor interaction since the estimate of error variance must be obtained from the sum of squares for three-factor interaction. These unreplicated three-way experiments take place for several reasons. Some experiments are too expensive to conduct so that experimenters are often forced to limit the number of treatment combinations that can be studied in order to have adequate resources available to replicate the treatment combinations under study. Sometimes these single-observation experiments occur by accident. That is, the researcher thought the experiment was replicated while, in reality, the replicates were really subsamples. How are such data to be analysed?

For three-way tables, analyses become complicated because one has to consider one more dimension than for two-way tables. For instance, the matrix theory, for two-way tables, can no longer be applied directly to three-way tables. In spite of the complexity of the subject, it has been suggested that the extensions of these ideas to deal with three-way interactions are also feasible. In some contexts, it makes sense to treat these interactions as if they were layers of two-way interactions and apply methods which exploit this viewpoint. More generally, an extended form of the singular value decomposition based on the work of Tucker (1963), Kroonenberg and De Leeuw (1980), and Harshman (1970) can be applied to these interactions although the method is not as simple as in the two-way case.

Three-mode principal component analysis model as introduced by Tucker (1963) and developed by Kroonenberg and De Leeuw (1980) results in what is referred to as Tucker3 model. The model decomposes the observation y as

$$y_{ijk} = \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} c_{pqr}\, g_{ip}\, h_{jq}\, e_{kr} + \epsilon_{ijk}, \qquad \begin{array}{l} i = 1, \ldots, l \quad j = 1, \ldots, m, \quad k = 1, \ldots, n \\ s \le l, \quad t \le m, \quad u \le n, \end{array}$$

which is a generalization of the well known machinery of data analysis associated with the singular value decomposition of two-way matrices. The model represents complex three-way interdependencies in a large data set ($l \times m \times n$) in terms of simpler interdependencies among the three sets of components as described by elements of a small three-way array ($s \times t \times u$) of a component interactions called the *core*. Notice that the term "three-mode" and "three-way" will be used interchangeably in this dissertation as they have similar meaning.

In this research, the three-mode principal component analysis model will be applied to decompose significant interaction in a three-factor experiment with no replication by modifying the usual three-way analysis of variance model as

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} c_{pqr}\, g_{ip}\, h_{jq}\, e_{kr} + \epsilon_{ijk}.$$

To differentiate interaction from random error, Chapter 3 presents three different approaches: randomization tests, likelihood ratio test, and Monte Carlo methods.

As a different line of approach, a graphical procedure has been proposed as a model diagnostic approach which may be viewed as a generalization to the graphical method for two-way data matrices. In particular, joint plot (Kroonenberg, 1983), as a graphical display of data arrays, is proposed as a data analytic tool for diagnosing the type of model to fit the data. Each entry of the data array is represented by a pair of vectors coming from the principal component scores. The model diagnosis proceeds by examining these vectors for collinearity in a two-dimensional plane or coplanarity in a three-dimensional Euclidean space. The relationship between various special forms of the general model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk}$$

and the geometrical configurations observed from the joint plots for three-way data arrays, allow one to diagnose the type of model to fit the data.

The primary application is to data from three-factor experiments with one observation per cell. These results apply to completely randomized three-factor experiments with no replication or to randomized block experiments with a single blocking factor and two crossed treatment factors (cf. remarks in Chapter 3).

This dissertation is organized as follows: Chapter 2 presents background information on analysis of a nonreplicated two-way table and three-mode principal component analysis (PCA). Chapter 3 and 4 describe feasible applications of three-mode PCA in three-factor fixed effects studies. Chapter 3 discusses procedures that can be applied to detect and to separate interactions from random error in a three-way table with one observation per cell. Chapter 4 illustrates a use of a graphical procedure, joint plots, which can be obtained in the process of three-way principal component analysis. Chapter 5 makes concluding remarks and suggests some areas for further research. Finally, a SAS program is given in Chapter 6.

CHAPTER 2. BACKGROUND

2.1 ANALYSIS OF NONREPLICATED TWO-WAY TABLE

In this section, the model and the analysis of nonreplicated two-way tables will be presented. When no interaction is presented, the hypothesis of no main effects is tested by using the ratios of the mean squares associated with the main effects to the error mean square. But when the interaction between the main effects is present, these tests are no longer valid. This has led to quite a bit of interest in the structure of interaction and the effect of interaction on the usual tests for main effects (Krishnaiah and Yochmowitz, 1980).

To state the problem formally, consider a general additive model with one observation y_{ij} per cell. A linear model associated with this is

$$y_{ij} = \mu + \alpha_i + \beta_j + \eta_{ij} + \epsilon_{ij}, \quad i = 1, \ldots, l \quad j = 1, \ldots, m, \tag{2.1}$$

where μ is the overall mean, α_i is the effect of the ith level of factor A, β_j is the effect of the jth level of factor B, η_{ij} is the effect of the interaction for the ith level of factor A and jth level of factor B, and ϵ_{ij} is error associated with that same combination and the ϵ_{ij}'s are assumed to be independently and identically distributed normal variates. This equation will hold for any two-way classification.

If there is no interaction, then η_{ij} will be equal to zero for all i and j. In that case, this will be an additive model and η_{ij} will drop out of (2.1). If there is an interaction, then two or more of the η_{ij} are unequal to zero and the model is no longer additive. In the analysis of interactions, the estimates $\hat{\mu}$, $\hat{\alpha}_i$, $\hat{\beta}_j$, are found for μ, α_i, β_j and residuals are calculated as $d_{ij} = y_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j$. The main interest is then in using d_{ij} to estimate η_{ij} of the model (2.1).

A common practice is to assume that the model is multiplicative. Tukey's (1949) model, also called the concurrent model is

$$y_{ij} = \mu + \alpha_i + \beta_j + \theta\alpha_i\beta_j + \epsilon_{ij}. \tag{2.2}$$

That is, it is assumed that the interaction term η_{ij} in the model (2.1) is a scalar multiple of the product of the row and column main effects.

Mandel (1961) generalized Tukey's model and proposed a linear model of the form

$$y_{ij} = \mu + \alpha_i + (1 + \delta\alpha_i^*)\beta_j + \epsilon_{ij}, \tag{2.3}$$

where the interaction is a linear function of β_j and called rows-linear model, and

$$y_{ij} = \mu + \beta_j + (1 + \xi\beta_j^*)\alpha_i + \epsilon_{ij}, \tag{2.4}$$

where the interaction is a linear function of the α_i and called the columns-linear model.

Bradu and Gabriel (1978) advocated the use of biplots on the matrix of residuals D, where

$$D = [d_{ij}] = [y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}]. \tag{2.5}$$

If the points or the endpoints of the vectors representing the columns (β_j) are collinear, it would suggest (2.3). If the points representing the rows (α_i) are collinear, it would suggest (2.4). If both sets were collinear, it would suggest the concurrent model (2.2). The additive model would be indicated by both sets being collinear but at right angles

to each other.

Applying concepts of principal component analysis for these models, Gollob (1968) and Mandel (1971) proposed another multiplicative model which partitions the interaction term, η_{ij}, into a sum of multiplicative functions of i and j

$$y_{ij} = \mu + \alpha_i + \beta_j + \sum_{p=1}^{r} \theta_p u_{pi} v_{pj} + \epsilon_{ij} \qquad (2.6)$$

where the ϵ_{ij}'s are assumed to be independently and identically distributed normal variates, and $r = \text{rank}(D) \leq \min(l-1, m-1)$. In vector notation, the conditions $\underline{u}'_p \underline{u}_p = \underline{v}'_p \underline{v}_p = \delta_{pp'}$ (Kronecker's delta) hold and u_p and v_p need not be functions of the main effects. The estimates of the interaction term in (2.6) come from a singular value decomposition of the residual matrix D (2.5). There will be a maximum of r eigenvalues and the sum of these eigenvalues equals the interaction sum of squares, where $\theta_1 \geq \theta_2 \geq \ldots\ldots \geq \theta_r > 0$. The estimate $\hat{\theta}_p^2$ is the pth largest eigenvalue of D'D or DD', $\hat{\underline{u}}_p$ is the eigenvector corresponding to the pth eigenvalue of DD', and $\hat{\underline{v}}_p$ is the eigenvector corresponding to the pth eigenvalue of D'D.

Gollob (1968) called this model (2.6) FANOVA, being a combination of factor analysis (principal component analysis, actually) and the analysis of variance. The model incorporates the benefits of data reduction from principal component analysis decomposition of residuals and the ease of interpretation permitted by the analysis of variance. The interaction sum of squares was partitioned into r terms $\hat{\theta}_1^2$, ..., $\hat{\theta}_r^2$. Generally, the last few roots are pooled as an error term and the others are tested by means of an F-tests. Since the roots are not independent and are not distributed as central chi-squares, these tests are not independent. In a slightly different approach, Mandel (1969, 1971) suggested adjusting the degrees of freedom. These degrees of freedom are based on the expected values of the eigenvalues, which are tabulated, using simulation, along with their standard deviations for the first few eigenvalues.

Johnson and Graybill (1972) examined a special case of model (2.6), namely, the presence of a single multiplicative component. For this case, they showed that the likelihood ratio test statistic (LRTS) of the first eigenvalue H_0: $\theta_1 = 0$ vs H_1: $\theta_1 \neq 0$ could be tested by using the statistic

$$\frac{\hat{\theta}_1^2}{\sum_{p=1}^{r} \hat{\theta}_p^2}.$$

The exact null distribution of this statistic has been derived when $(l-m-1)/2$ is an integer by Schuurmann, Krishnaiah, and Chattopadhyah (1973). Corsten and Van Eijnsbergen (1972) showed that the group of components for equality, that is, for simultaneously testing H_0: $\theta_1 = ... = \theta_c = 0$, could be tested by using the likelihood ratio test statistic

$$\frac{\sum_{p=1}^{c} \hat{\theta}_p^2}{\sum_{p=1}^{r} \hat{\theta}_p^2}$$

where $c < r \leq \min(l-1,\ m-1)$. Hegemann and Johnson (1976) derived the likelihood ratio test statistic for testing the hypothesis H_0: $\theta_2 = 0$ for the second component of the form

$$\frac{\hat{\theta}_2^2}{\sum_{p=2}^{r} \hat{\theta}_p^2}$$

and included some tables for its use. Finally, Yochmowitz and Cornell (1978) showed that this can be generalized so that the test statistic for the kth root, H_0: $\theta_k = 0$, is

$$\frac{\hat{\theta}_k^2}{\sum_{p=k}^{r} \hat{\theta}_p^2}.$$

Once a multiplicative component is found to differ significantly from zero, then the cause of the interaction may be investigated. Snee (1982) studied the coefficients in the fitted model to distinguish between row- or column-related nonhomogeneous variance and interaction between the row and column vectors.

The methods proposed here are for nonreplicated two-way tables. In contrast to two-way tables, it is somewhat surprising to note that not much work has been presented for

the unreplicated three-way tables. The following section describes three-mode principal component analysis. In chapter 3, it will be shown that this is a useful technique to analyse the interactions in three-way tables with only one observation per cell.

2.2 THREE-MODE PRINCIPAL COMPONENT ANALYSIS

This section describes three-mode principal component analysis, a method for multivariate data, based on the book by Kroonenberg (1983). In later chapters, this method will be applied to three-way analysis of variance models with a single observation per cell.

2.2.1 Introduction

Investigating relationships in multivariate data is a necessary but often difficult task. Such data needs to be condensed and summarized in a rather simple form that is to be more easily comprehended. Among the most popular methods to achieve such parsimonious summarization is principal component analysis (PCA), which looks for a few linear combinations that can summarize the data, keeping in the process as much information as possible.

The central idea of principal component analysis is to reduce the dimensionality of a data set which consists of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set. This is achieved by transforming to a new set of variables, the principal components, which are uncorrelated and which are ordered, so that the first *few* retain most of the variation present in *all* of

the original variables.

Suppose we have data on a population of 11 distinct species measured at 12 different sites once every time period over 6 time periods. The data can be assembled into a cubical array in which the entry y_{ijk} is the population density of the species i at site j at time k. Since each entry in the array is subject to the three ways or *modes* of classification - by species, site, and time - this would be considered *three-mode* data.

Until not long ago, a researcher who wanted to perform component analysis of such data would have to force the data into the standard two-way form by combining two of the classifications, such as site and time, perhaps by averaging successive slices of the array, leaving species as the other dimension and reducing the problem into two dimensions. Other times, such data was analysed separately at each time period, i.e., one might carried out a separate PCA of the standard two-way matrix at each time period instead of performing the analysis simultaneously. But these ignore potentially important three-way interdependencies among the observations.

More recently, some techniques have been developed to obtain direct solutions for three-mode data sets, and the technique is generally referred to as Three-Mode (Three-Way) Principal Component Analysis. The procedure was first formulated by Tucker (1963, 1966), who generalized the decomposition of the two-way component analysis to the three-way array. The general aim of the three-mode analysis is to fit a model to the data with a low dimensional representation so that the basic underlying structure can be more readily determined and interpreted. This is achieved by computing principal component analysis for two or three modes.

Principal component analysis is directly related to a more general decomposition method based on the singular value decomposition (SVD). In doing a principal component analysis for a given data matrix, the importance of the SVD is threefold. It provides a computationally efficient method of actually finding principal components; it supplies additional insight into what a PCA actually does; and it gives useful means, both graphical and algebraic, of representing the results of principal component analysis. The SVD of an arbitrary l x m matrix Y can be written as

$$Y = U\Lambda V' \tag{2.7}$$

where $U'U = V'V = I_r$, Λ is $r \times r$ diagonal, U is l x r, and V is m x r, where $r \leq \min(l, m)$. In summation notation, this is

$$y_{ij} = \sum_{p=1}^{r} \theta_p u_{ip} v_{jp}.$$

The decomposition is not unique in the sense that any permutation of columns of U is allowed, provided that the inverse permutation is applied to Λ and V. Here we will assume that all θ_p are different, and they are arranged in descending order. With these side conditions the SVD of Y is uniquely determined.

U and V are the eigenvector matrices of YY' and $Y'Y$ respectively, and the singular values θ_p are the square roots of the eigenvalues. At most $\min(l, m)$ of the θ_p can be nonnegative. An important property of the SVD is that the best least squares approximation of Y of rank $r < \min(l, m)$, is equal to $Y_r = U_r \Lambda_r V_r'$ with U_r and V_r containing the first r columns of U and V respectively, and with Λ_r the diagonal matrix with the largest r singular values. The diagonality of Λ implies that the pth component of U is exclusively linked with the pth component of V, i.e., there is no relation between the pth component of U and p'th component of V.

From a methodological point of view, three-mode PCA is a generalization of standard PCA, or rather, of the SVD. However, not all properties of the two-mode SVD carry over to all three-way models, and no three-way model has all the properties analogous to the two-way SVD. The difficulties in the extension of the SVD to the case of three-way arrays are mainly due to the many mathematical differences between three-way arrays and two-way arrays with respect to decomposition and rank as discussed in Kruskal (1989).

There are several decompositions of three-way data which claim to be the three-way generalization of the SVD. The general principal component model (Tucker3 model), which forms the basic working model for this research, will be described in the next section.

2.2.2 The Model

Tucker (1966) developed three-mode principal component analysis by generalizing standard principal component analysis, or rather, the singular value decomposition of two-way data to three-way data sets. A $l \times m \times n$ three-way table Y is defined as the collection of elements

$$\{y_{ijk} |\ i = 1, \cdots, l;\ j = 1, \cdots, m;\ \ k = 1, \cdots, n\}.$$

The elements are placed in the three-dimensional block such that indices i, j, k run along the vertical axis, the horizontal axis, and the *depth* axis, respectively. The word *mode* is used to indicate a collection of indices by which the data can be classified. Here only real matrices will be used and the notations for these are: $\Re^{l \times m \times n}$ for the class of all $l \times m \times n$ three-mode arrays, and $K^{m \times n}$ for the class of $m \times n$ columnwise orthonormal matrices. While many models have been suggested for the analysis of three-mode data, the most general is the Tucker3 model (Kroonenberg, 1983). Using the above definitions, the Tucker3 model, which is the factorization of the three-mode data $Y = [y_{ijk}] \in \Re^{l \times m \times n}$, can be written as

$$y_{ijk} = \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} c_{pqr}\, g_{ip}\, h_{jq}\, e_{kr}, \tag{2.8}$$

where the coefficients g_{ip}, h_{jq}, and e_{kr} are the elements of the component matrices $G \in K^{l \times s}$, $H \in K^{m \times t}$, and $E \in K^{n \times u}$ respectively, and s, t, u are the number of components of the first, second, and third mode. The c_{pqr} are the elements of the so-called three-mode core array $C \in \Re^{s \times t \times u}$ and each of these elements represents a unique combination of categories of the components. A matrix formulation of the Tucker3 model is

$$Y = GC(E \otimes H)' \tag{2.9}$$

where $Y \in \Re^{l \times mn}$, $C \in \Re^{s \times tu}$ are now ordinary two-mode matrices, each of which is obtained by concatenating frontal slices in a row, and $(E \otimes H) = [e_{kr}H]$ denotes the Kronecker product of matrices.

To explore the relation between components from the various sets, consider a two-way table. The squared singular values, θ_p^2, of Λ in $Y = U\Lambda V'$ (2.7), express the amount of variance accounted for by the pth component of U and V. For a three-way table, if G, H, and E are columnwise orthonormal, the squared core elements, c_{pqr}^2 in (2.8), does the same for the combination of the three components. Thus, c_{pqr}^2 is the amount of variance, or sum of squares, jointly accounted for by the pth, qth, and rth component of G, H, and E respectively.

The next section presents the main theory for the method of alternating least squares, due to Kroonenberg and De Leeuw (1980), for fitting the Tucker's three-mode principal component model (Tucker3 model). It also describes the development of the TUCKALS3 algorithm which estimates the parameters in the three-mode principal component analysis model and thus obtain the solutions to (2.10) below.

2.2.3. Method and Algorithm

Loss Function

As mentioned earlier, if all the principal components are computed, thus, $s = l$, $t = m$, and $u = n$, then the data array could be decomposed exactly into its components. However, in practical applications, one is just interested only in the first few principal components for each mode. This precludes finding an exact factorization of Y in G, H, E, and C. One therefore has to be satisfied with an approximation, $\hat{Y} = GC(E' \otimes H')$,

that is, finding G, H, E, and C such that s, t, and u are as small as possible yet the difference between the model and the data is minimal according to some loss function. The least squares loss function corresponding to (2.8) has the form

$$f(G,H,E,C) = ||Y-\hat{Y}||^2 = || Y-GC(E' \otimes H')||^2 . \tag{2.10}$$

where ||.|| denotes the Euclidean norm. The minimization has to be carried out under the restrictions of the model, i.e., G, H, and E must be columnwise orthonormal matrices. There always exist some G, H, E and C such that the loss function f in (2.10) attains a global minimum (cf. Page 73 of Kroonenberg and De Leeuw (1980)). This is shown in two steps. First, for given G, H, and E, find a unique $\hat{C}$ which can be expressed in terms of G, H, E, and Y, and secondly, the resulting $\hat{C}$ minimizes the loss function f (2.10). In this way the problem is reduced to minimizing f over G, H, and E. Thus for fixed G, H, and E, f attains its minimum for this $\hat{C}$, where,

$$\hat{C} = G'Y(E \otimes H) \tag{2.11}$$

and its elements

$$\hat{c}_{pqr} = \sum_{i=1}^{s} \sum_{j=1}^{t} \sum_{k=1}^{u} g_{ip}\, h_{jq}\, e_{kr}\, y_{ijk}. \tag{2.12}$$

The elements of the core C are referred as generalized singular values.

To find $\hat{Y}$ which minimizes the loss function f in (2.10), substitute (2.11) into (2.10), and call the rewritten function g

$$\begin{aligned} g(G,H,E) = ||Y-\hat{Y}||^2 &= ||Y-GG'Y(H \otimes E)(H' \otimes E')||^2 \\ &= ||Y-GG'Y(HH' \otimes EE')||^2 . \end{aligned} \tag{2.13}$$

The domain S of the function g is

$$S = \{s \mid s = (G,H,E),\ G \in K^{lxs},\ H \in K^{mxt},\ E \in K^{nxu}\}$$

which is a compact subset in a finite-dimensional Euclidean space. Using the fact that g

is a bounded continuous function on S ($0 \leq g \leq ||Y||^2$), it is concluded that there exists a point $s = (\hat{G},\hat{H},\hat{E})$ in S, such that g attains its minimum. In other words the minimization problem always has a solution.

In order to reduce the minimization problem (2.10) to some handier form, the problem is converted to a maximization problem. Rewriting the equation (2.13) using the traces instead of norms, it is found that

$$\begin{aligned} g(G,H,E) &= \mathrm{tr}(Y-\hat{Y})(Y-\hat{Y})' \\ &= \mathrm{tr}(Y-GG'Y(HH'\otimes EE'))(Y-GG'Y(HH'\otimes EE'))' \\ &= \mathrm{tr}YY'-2\mathrm{tr}GG'Y(HH'\otimes EE')Y'+\mathrm{tr}G'Y(HH'\otimes EE')Y'G \\ &= \mathrm{tr}YY'-\mathrm{tr}G'Y(HH'\otimes EE')Y'G \end{aligned}$$

or

$$SS(\text{Residual}) = SS(\text{Total})-SS(\text{Fit}).$$

Clearly the minimization of the residual sum of squares g(G,H,C) comes down to the maximization of the fitted sum of squares:

$$SS(\text{Fit}) = p(G,H,E) = \mathrm{tr}G'Y(HH'\otimes EE')Y'G \tag{2.14}$$

still under the constraints that G, H, and E are columnwise orthonormal. So far, H and E are placed in the Kronecker product term, but we could equally well have done so with G and E, or G and H. Such substitutions cause only a change in form, but not in the model itself. The model is indifferent to such notational changes as can be clearly seen from (2.8).

The necessary and sufficient conditions for the existence and nature of an exact solution to the minimization problem (2.10) are stated in the theorems 1 and 2 in Kroonenberg and De Leeuw (1980). In essence, these theorems tell us that SS(Fit) (2.14) can be maximized by simultaneously solving the eigenvalue-eigenvector problems of P*, Q*, and R*:

$$P^*(H,E) = Y(HH'\otimes EE')Y' \quad \text{and } Y \in R^{l\times mn} \tag{2.15}$$

$$Q^*(E,G) = Y(EE'\otimes GG')Y' \quad \text{and } Y \in R^{m\times nl} \tag{2.16}$$

$$R^*(G,H) = Y(GG' \otimes HH')Y' \text{ and } Y \in R^{nxlm} \tag{2.17}$$

where

G, H, and E are the eigenmatrices corresponding to the eigenvalues of $P^*(H,E)$, $Q^*(E,G)$, and $R^*(G,H)$, respectively. (2.18)

Notice that although $\hat{c}_{pqr}$ is not a direct function of the eigenvalues, it is related to the eigenvalues as follows. First, we note that

$$\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} y_{ijk}^2 = \sum_{p=1}^{l} \lambda_p^* = \sum_{q=1}^{m} \mu_q^* = \sum_{r=1}^{n} \nu_r^*, \tag{2.19}$$

where,

$\lambda_1^* > \cdots > \lambda_l^*$ are the eigenvalues of $P^*(H,E)$

$\mu_1^* > \cdots > \mu_m^*$ are the eigenvalues of $Q^*(E,G)$

$\nu_1^* > \cdots > \nu_n^*$ are the eigenvalues of $R^*(G,H)$

Also from (2.8), we obtain

$$\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} y_{ijk}^2 = \sum_{p=1}^{s}\sum_{q=1}^{t}\sum_{r=1}^{u} \hat{c}_{pqr}^2 \tag{2.20}$$

if $s = l$, $t = m$, and $u = n$ (cf. Appendix A). Now, by combining the equations in (2.19) and (2.20),

$$\sum_{p=1}^{s}\sum_{q=1}^{t}\sum_{r=1}^{u} \hat{c}_{pqr}^2 = \sum_{p=1}^{s} \hat{\lambda}_p^* = \sum_{q=1}^{t} \hat{\mu}_q^* = \sum_{r=1}^{u} \hat{\nu}_r^*. \tag{2.21}$$

Furthermore, we obtain the following relations between the eigenvalues of the cross product matrices of Y and the core elements:

$$\hat{\lambda}_p^* = \sum_{q=1}^{t} \sum_{r=1}^{u} \hat{c}_{pqr}^2, \tag{2.22}$$

$$\hat{\mu}_q^* = \sum_{p=1}^{s} \sum_{r=1}^{u} \hat{c}_{pqr}^2, \tag{2.23}$$

$$\hat{\nu}_r^* = \sum_{p=1}^{s} \sum_{q=1}^{t} \hat{c}_{pqr}^2. \tag{2.24}$$

Notice that, in P*(H,E), Q*(E,G), and R*(G,H), the first s, t, and u eigenvectors are required for (2.15), (2.16), and (2.17) respectively. Thus the component matrices $\hat{G}$, $\hat{H}$, and $\hat{E}$ of a solution are each nothing but the eigenvectors corresponding to the largest eigenvalues of suitably constructed cross-products of the data matrix Y and the other two components matrices. These solutions, however, are not unique and any orthonormal transformation of G and/or H, and/or E will provide a solution as well. Not surprisingly this can not be solved analytically, but only iteratively.

Alternating Least Squares (ALS) Algorithm

Obviously one would like to construct an algorithm for the maximization of the fitted sum of squares, p, in (2.14), that converges to a global maximum. Unfortunately p is the cross-product term of a high degree multivariate polynomial, and in general it is not possible to prove that methods to solve such nonlinear problems attain a global optimum.

The method to be described utilizes alternating least squares, an iterative procedure. The essential feature of the ALS approach is that in solving optimization problems with more than one set of parameters, each set is estimated in turn by applying conditional least squares procedures holding the other sets fixed. After all sets have been estimated once, the procedure is repeated again and again until convergence.

In order to see how the alternating least squares approach can be applied in the present context, recall (2.13), $g(G,H,E) = ||Y - GG'Y(HH' \otimes EE')||^2$. The sets of parameters here are G, H, and E, as C can be derived from these three as in (2.11). Minimizing g over G

holding H and E fixed is identical to solving one conditional least squares problem, minimizing over H holding E and G fixed, and minimizing over E with G and H fixed are the two others. Although the goal here is to maximize the SS(Fit), p in (2.14), the problem is still an ALS one.

A rough outline for an algorithm is given as follows:

step1. choose an arbitrary H_0 and E_0 and maximize over G to get a new G_1,
step2. maximize subsequently over H with the just computed G_1 and E_0 fixed to get a new H_1, and
step3. finally maximize $p = SS(Fit)$ over E with G_1 and H_1 fixed yielding a new E_1.
Iterate this procedure until convergence.

With the $\hat{G}$, $\hat{H}$, and $\hat{E}$ obtained from the procedure, $\hat{C}$ is now given by $\hat{G}'Y(\hat{E} \otimes \hat{H})$ as in (2.11). The maximizations are essentially identical to searching for eigenvectors and eigenvalues of matrices of the order l, m, and n respectively. As the orders of P, Q, R are typically much larger than the number of components or eigenvectors described, Kroonenberg and De Leeuw (1980) chose an eigenvalue-eigenvector technique, known as Bauer-Rutishauser simultaneous iteration method (Rutishauser, 1969), which is particularly efficient in this situation. The algorithm is based on an eigenvalue-eigenvector technique. The maximization of p(G,H,E) in (2.14) consists of an iteration process in which, at each step, three eigenvalue-eigenvector problems have to be solved. For further discussion on this iteration, we refer to Chapter 4 of Kroonenberg (1983).

CHAPTER 3. TESTS FOR COMPONENTS OF INTERACTION IN THREE-WAY TABLES

3.1 INTRODUCTION

Earlier the three-mode principal component model was discussed as a useful technique to analyse three-way data. This chapter discusses how to use the three-mode principal component analysis in the study of interactions in a three-way ANOVA table with only one observation per cell.

As a starting point, one can take the fully general three-way classification model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijk},$$

$$i = 1, \ldots, l \quad j = 1, \ldots, m, \quad k = 1, \ldots, n, \tag{3.1}$$

where μ is an overall mean; α_i, β_j, and γ_k are the ith, jth, and kth levels of the main effects of three factors, say A, B, and C respectively; $(\alpha\beta)_{ij}$ is the effect of the interaction for the ith level of the factor A and the jth level of the factor B, and similarly $(\alpha\gamma)_{ik}$ and $(\beta\gamma)_{jk}$ are the corresponding two-way interaction terms; $(\alpha\beta\gamma)_{ijk}$ is the three-way interaction term and the ϵ_{ijk} term represents random error and the ϵ's are

assumed to be independently and identically distributed normal variates with mean 0 and variance σ^2, i.e., $\epsilon_{ijk} \sim \text{NID}(0, \sigma^2)$. To give the parameters a unique interpretation, usual constraints are imposed, i.e., all of the main effect and interaction terms are constrained to sum to zero across each of their subscripts.

Now consider an experiment where the basic treatment structure is three-way and there are no independent replications of the three-way treatment combinations. Standard ANOVA models for such data, which include all possible main effect and interaction terms, no longer work because the number of parameters in the model equals the number of data points. In other words, the usual methods of statistical analysis no longer apply because there are no independent replications from which to estimate the experimental error variance σ^2. Having a good estimate of σ^2 is important when the major objective of the experiment is confirmatory. However, if the major objective of the experiment is exploratory, it is often more desirable to study different treatment combinations, each performed once, rather than a few treatment combinations each replicated many times. This chapter is devoted to the methods that can help to extract the relevant information in three-factor experiments that are not replicated.

One way to handle this situation is to omit the three-way interactions, possibly even some of the two-way interactions, to simplify the model. In some cases, there is no reason to believe that these interactions are null. The assumption of no interactions should not be made without some justification for it being true. But many experimenters are more than willing to assume the factors, say, A, B, and C do not interact, especially when such an assumption enables them to obtain some test statistics. What can be done? What can an experimenter do when three-factor interaction is suspected to be present in the data?

One natural way of generalizing the techniques in section 2.1 on three-way treatment structure is to look at the experiment as though it is two-way. For the three-way model this can be done by letting all possible combinations for two of the factors in the three-way represent one of the factors in the two-way and letting the levels of the third factor in the three-way represent the second factor in the two-way. That is, if one had a three-way treatment structure with factors A, B, and C, one could consider the experiment as a two-way with factors A*B and C, A*C and B, or A and B*C. It can be easily shown

that if there is no two-way interaction between the levels of A*B and C, then there can be no three-way interaction between the levels of A, B, and C. However, if there is interaction between the levels A*B and C, there may or may not be a three-way interaction (Milliken and Johnson, 1989). In this case, it would be wise to look for interaction between the levels of A*C and B and/or between the levels of B*C and A.

For a three-way $l \times m \times n$ array of data Y, to diagnose three-factor interaction in the above sense the data array must be reduced to matrix form. This may be done by crossing two of the factors to comprise the rows and isolating the remaining factor levels in the column. Thus it would become, for example, an $lm \times n$ data matrix. There are six such matrix reductions available. Crossing the levels of two factors to reduce three-way data to each of the six matrix forms results in asymmetric treatment of the three two-factor interactions that has to be considered in the decomposition in (3.1). So such a reduction method may be applicable for a special type of data. For an example, see Chapter 6 of Milliken and Johnson (1989).

Another reduction method was introduced by Kettenring (1983). In order to check if there is two-way interaction between the factors A and B in three-way data, he reduced the $l \times m \times n$ array to the $l \times m$ matrix $[d_{ij}] = [y_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}]$, i.e., an average is being taken with respect to the third subscript. Once the three-way data array is reduced to a two-way matrix then the methods available for two-way interactions can be applied.

These methods are essentially forcing the three-way array into the standard matrix form so that the known methods for two-way data can be applied. But these ignore potentially important three-way interdependencies among the observations. One could not recommend such procedures to analyse three-way interaction if there is a way to analyse all three factors simultaneously since one can gain much more insight into the data structure.

While few specific solutions have been developed for this more difficult situation, one possibility is to extend the concurrent model to cover all the interaction terms involved as in Kester (1979). Her model uses the same constant of proportionality for each of the multiplicative terms representing two-way interactions and its square for the three-way multiplicative terms. Kettenring (1983) used the modified parallel factor analysis model

to decompose the three-factor interaction term in a nonreplicated three-way table. He replaced $g_{ip}h_{jp}e_{kp}$ in (2.12) with $c_{ppp} = (\sum g_{ip}^2 \sum h_{jp}^2 \sum e_{kp}^2)^{1/2}$, $g_{ip}^* = \frac{g_{ip}}{\sqrt{\sum g_{ip}^2}}$, $h_{jp}^* = \frac{h_{jp}}{\sqrt{\sum h_{jp}^2}}$, and $e_{kp}^* = \frac{e_{kp}}{\sqrt{\sum e_{kp}^2}}$ for all p (which makes this model appear more like the SVD), applying this form, to the three-way sample interactions.

This chapter is organized in the following manner: Section 3.2 proposes a new analysis of variance model for a data with a three-way cross-classification treatment structure with only one observation per treatment combination by applying three-mode principal component analysis to the residuals. Sections 3.3 and 3.4 discuss two different approaches for testing the null hypothesis of no three factor interactions: randomization tests and likelihood ratio test. Section 3.5 shows a third approach which applies Monte Carlo methods to separate interaction from random error.

3.2 THREE-WAY PCA IN ANOVA MODELS

Using the common ANOVA notation, the nonreplicated three-way model is usually written as

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \epsilon_{ijk} \qquad (3.2)$$

with the standard meaning and constraints on parameters. This model is like model (3.1) but the $(\alpha\beta\gamma)_{ijk}$ term is missing. As stated before, because there is only one observation per each treatment combination, one often assumes that there is no three factor interaction and this term is used to represent ϵ_{ijk} which is considered as the *pure* random error.

The more general model for this nonreplicated three-way table can be written as

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \eta_{ijk} + \epsilon^*_{ijk} \qquad (3.3)$$

In this model, ϵ_{ijk} in (3.2) is no longer considered as just random error. However a part of ϵ_{ijk} may be just random error so the term is written as a sum of two separate quantities

$$\epsilon_{ijk} = \eta_{ijk} + \epsilon^*_{ijk},$$

where η_{ijk} is referred to as the effect of interaction between the ith, jth, and kth levels of the three factors, and ϵ^*_{ijk} is the experimental error associated with that same combination and the ϵ's are assumed to be independently and identically distributed normal variates. As usual, in addition to the standard constraints, the conditions, $\sum_i \eta_{ijk} = \sum_j \eta_{ijk} = \sum_k \eta_{ijk} = 0$ are imposed.

The goal is then to separate or partition the three-factor interaction sum of squares into a portion related to the interactions and a portion associated with random error. The decomposition of the interaction term η_{ijk} into the sum of multiplicative functions of i, j, and k using the three-mode principal component analysis is represented by

$$\eta_{ijk} = \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} c_{pqr}\, g_{ip}\, h_{jq}\, e_{kr} \qquad (3.4)$$

where s, t, and u are the number of the components of the first, second, and third factor of the residual array below (3.6). In equation (3.4), the interaction parameters η_{ijk} are expressed as the sum of several multiplicative contrasts such that each contrast is orthogonal to all previous contrasts and accounts for a maximum of the variance of the η_{ijk}.

In practice, the partitioning of the interaction into multiplicative terms is carried out only partially, i.e., only a few multiplicative terms are retained and the remaining terms are pooled together and considered as experimental error. The values of s, t, and u are much smaller than l, m, and n, and these are often one or two. If $s = l$, $t = m$, $u = n$, then

η_{ijk} is decomposed exactly into its components and so

$$\eta_{ijk} = \epsilon_{ijk} = \sum_{p=1}^{l} \sum_{q=1}^{m} \sum_{r=1}^{n} c_{pqr}\, g_{ip}\, h_{jq}\, e_{kr}.$$

The usual three-way fixed effects ANOVA model, augmented by a three-way principal component representation of three-way interaction terms (3.4), gives the following multiplicative model,

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} c_{pqr}\, g_{ip}\, h_{jq}\, e_{kr} + \epsilon^*_{ijk}. \qquad (3.5)$$

The established parameters of the model (3.5) are obtained in two steps. First, the standard ANOVA model (3.2) is fitted to the given data. Second, the residuals from the first step are decomposed using three-mode PCA on the residual array $D \in \Re^{lxmxn}$, where

$$\begin{aligned} D = [d_{ijk}] &= [y_{ijk} - \hat{\alpha}_i - \hat{\beta}_j - \hat{\gamma}_k - (\widehat{\alpha\beta})_{ij} - (\widehat{\alpha\gamma})_{ik} - (\widehat{\beta\gamma})_{jk}] \\ &= [y_{ijk} - \bar{y}_{ij.} - \bar{y}_{i.k} - \bar{y}_{.jk} + \bar{y}_{i..} + \bar{y}_{.j.} + \bar{y}_{..k} - \bar{y}_{...}]. \end{aligned} \qquad (3.6)$$

Using a three-mode representation with complete decomposition, i.e., $s = l$, $t = m$, $u = n$, the residuals can be written as

$$d_{ijk} = \sum_{p=1}^{l} \sum_{q=1}^{m} \sum_{r=1}^{n} \hat{c}_{pqr}\, \hat{g}_{ip}\, \hat{h}_{jq}\, \hat{e}_{kr}. \qquad (3.7)$$

But in practice, only a few terms of these are retained and the remaining terms are pooled and considered as error:

$$d_{ijk} = \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} \hat{c}_{pqr}\, \hat{g}_{ip}\, \hat{h}_{jq}\, \hat{e}_{kr} + (\text{error})_{ijk}. \qquad (3.8)$$

This produces the estimates $\hat{c}_{pqr}$, $\hat{g}_{ip}$, $\hat{h}_{jq}$, and $\hat{e}_{kr}$, where the generalized singular values $\hat{c}_{pqr}$ are the elements of the core array, and $\hat{g}_{ip}$, $\hat{h}_{jq}$, and $\hat{e}_{kr}$ are the elements of the columnwise orthonormal component matrices G, H, and E, where these are the eigenvector matrices of the cross products of the residuals reduced by the components of the other factors (cf. (3.22) to (3.24)). These estimates have the usual constraints,

$$\sum_{i=1}^{l} \hat{g}_{ip} = \sum_{j=1}^{m} \hat{h}_{jq} = \sum_{k=1}^{n} \hat{e}_{kr} = 0 \ , \quad \forall \ p, \ q, \ r.$$

$$\sum_{i=1}^{l} \hat{g}_{ip}^2 = \sum_{j=1}^{m} \hat{h}_{jq}^2 = \sum_{k=1}^{n} \hat{e}_{kr}^2 = 1 \ , \quad \forall \ p, \ q, \ r.$$

$$\sum_{i=1}^{l} \hat{g}_{ip}\hat{g}_{ip'} = \sum_{j=1}^{m} \hat{h}_{jq}\hat{h}_{jq'} = \sum_{k=1}^{n} \hat{e}_{kr}\hat{e}_{kr'} = 0 \ , \quad \forall \ p \neq p' \ , \ q \neq q', \ r \neq r'. \tag{3.9}$$

The following three sections deal with issues related to the problem of how to detect and separate the interaction from random error. First, section 3.3 discusses randomization tests.

3.3 RANDOMIZATION TESTS

This section shows how randomization tests can be applied in the area of testing for the structure of three-factor interaction term η_{ijk} (3.4). The analysis is applied on residuals in (3.6).

3.3.1 The idea of randomization tests

Randomization tests are used to determine whether the null hypothesis of randomness is reasonable, when the alternative hypothesis suggests that there will be a tendency for a certain type of pattern to appear in data. In particular, randomization tests are procedures for determining statistical significance directly from experimental data

without recourse to significance tables. Suppose a statistic, say S, is chosen to measure the extent to which data show the significance of three-factor interaction. The value of s_0 of S for the observed residuals is then computed with the distribution of S that is obtained by randomly reordering the residuals. The argument made is that if the null hypothesis is true then all possible orders for the data were equally likely to have occurred. The observed residuals order is then just one of the equally likely orders and s_0 should appear as a typical value from the randomization distribution of S. If this does not seem to be the case so that s_0 is significant then the null hypothesis is discredited to some extent and, by implication, the alternative hypothesis is considered more reasonable.

The significance level of s_0 or p-value is the proportion of data permutations providing as large a test statistic value as that associated with the experimental results. This can be interpreted in the same way as conventional tests of significance: if it is less than .05 then this provides some evidence that the null hypothesis is not true, if it is less than .01 then it provides stronger evidence that the null hypothesis is not true, and so on. In practice, it is more desirable to report the p-value, the smallest α at which the results are significant, itself, rather than providing a conclusion based on the conventional levels of significance.

There are two basic methods of permuting data by the use of a randomization test. One way is to use all possible data permutations in a set or subset to determine significance. In many cases, the large number of possible data permutations makes a complete enumeration extremely difficult if not impossible. An alternative way is to use random data permutations which uses a random sample of all possible data permutations to determine significance. It reduces a substantial number of permutations while it serves the same function as the earlier method. Instead of requiring several million or billion data permutations, the random data permutation method may be effective with as few as 500 data permutations. In this research, the randomization test will be performed using random data permutation instead of using the all possible configurations of the data.

The null hypothesis of the randomization test is the set of observations (i.e., residuals, in our case) associated with each assignment unit is independent of the assignment of

units to treatments. That is, the set of residuals $\{d_{ijk} | \ i = 1,\cdots,l, \ j = 1,\cdots,m, \ k = 1,\cdots,n\}$ were obtained regardless of the assignment of units to the treatments. In particular, for the model (3.5), the null hypothesis of no three-factor interaction of a randomization test is

$$H_0 : c^2_{pqr} = 0, \text{ for all } p = 1,\cdots s, \ q = 1,\cdots t, \ r = 1,\cdots u.$$

This implies that the model (3.2), with no three-factor interaction term η_{ijk}, as the null model.

3.3.2 The procedures of randomization tests

In this section, the procedure of the randomization test will be discussed. First the residuals $D \in \Re^{l \times m \times n}$ (3.6) are obtained by fitting the standard three-way ANOVA model (3.2) to the given data. To test whether the term η_{ijk} (3.4) belongs in the model (3.3), the obtained residuals are permuted repeatedly between all $l \times m \times n$ treatments. For each permutation of the residuals, a test statistics S is computed using three-mode PCA. This gives an empirical distribution of the test statistic S and it will be called as the randomization distribution as Manly (1990) did. The computed value s is then compared with the value s_0, from the observed residuals, for each permutation. Finally, the p-value or the significance level of s_0 is computed by determining the proportion of values that are as extreme or more extreme than this value s_0 in the randomization distribution.

If s_0 looks like a typical value from the randomization distribution then conclude that the assignment of units to the treatments that occurred in the residuals in reality seems to be random. In other words, conclude that there is no three-factor interaction and do not rejects H_0: $\eta_{ijk} = 0$ for all i, j, k (or equivalently, $c^2_{pqr} = 0$ for all p, q, r). On the other hand, if s_0 is unusually large then the data are unlikely to have arisen if the null model is true and it can be concluded that the alternative hypothesis H_1: $\eta_{ijk} \neq 0$, or

equivalently, $c^2_{pqr} > 0$ for at least one term in (3.4) is more plausible.

It is a question of whether s_0 is unusually large in comparison with the randomization distribution. If s_0 is in the bottom, say, 95 % part of the randomization distribution, the test result is not significant so no three-factor interaction terms should be retained in the model (3.2). If s_0 is among the values in the top 5% of the distribution then the result is significant so the model should include the term which represents three-factor term at the 5% level. A convenient way to summarize the randomization results involves calculating the p-value, the proportion of all the observed values of S that are greater than or equal to s_0. This is an estimated probability of such a large value, which is the significance level of s_0.

In summary, the steps for the randomization test is as follows:

1. Obtain the residuals (3.6) by fitting the standard three-way ANOVA model (3.2) to the given data.

2. Compute s_0, which is a test statistic value for the results given by the observed residuals.

3. Randomly permute the residuals and compute the test statistic value s for each set of permuted residuals.

4. Repeat step (3) a large number of times to find a sample of values from the distribution of S.

5. Count each time whenever $s \geq s_0$, i.e., a test statistic value is as large as the obtained value.

6. Compute the p-value:

$$\text{p-value} = \frac{\text{number of } \{s \geq s_0\} + 1}{\text{number of permutation} + 1}$$

That is, the p-value is the proportion of the test statistic values, including the obtained value, that are as large as the obtained test statistic value.

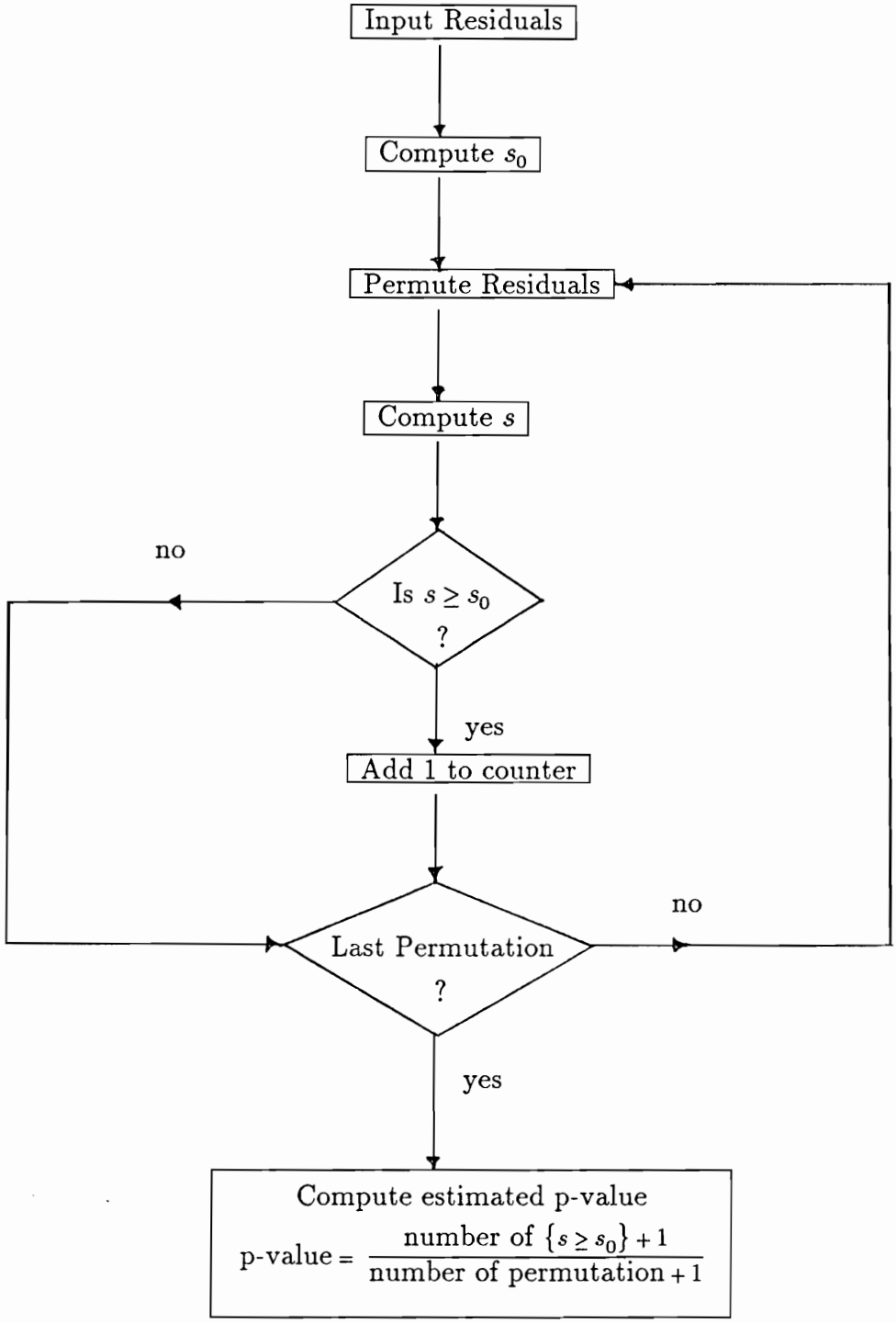


Figure 1. Flow chart for randomization tests

Figure 3.1 shows the basic structure of the above procedure.

3.3.3 Choice of test statistic

Before we discuss test statistics, some notations are introduced. For a given array $A = (a_{ijk})$, let $||A||^2$ denote the sum of squares of all elements of A. In particular,

$$||D||^2 = \sum_{i=1}^{l} \sum_{j=1}^{m} \sum_{k=1}^{n} d_{ijk}^2 = SS(\text{Residual}) \tag{3.10}$$

and this is the residual sum of squares of $D \in \Re^{l \times m \times n}$. When there is only one observation per treatment combination, this quantity represents the sum of squares for interaction between three factors. Furthermore, if $s = l$, $t = m$, and $u = n$, then using (3.7)

$$||D||^2 = ||C||^2 = \sum_{p=1}^{l} \sum_{q=1}^{m} \sum_{r=1}^{n} c_{pqr}^2, \tag{3.11}$$

where $C \in \Re^{l \times m \times n}$ (For the proof, see Appendix A). If $s < l$, $t < m$, and $u < n$, then

$$||C||^2 = \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} c_{pqr}^2 \tag{3.11a}$$

and this quantity represents the sum of squares of the fit of the residual array D, where $C \in \Re^{s \times t \times u}$.

Notice also that D can be decomposed and written in matrix form as

$$D = GC(E \otimes H)' \tag{3.12}$$

(cf. (2.9)). The matrix form $D \in \Re^{l \times mn}$ is obtained by connecting the frontal matrices (cf. (4.10)) of the array in a row. There are two more different ways to write the three-way array D into a matrix form: the form $D \in \Re^{m \times ln}$ is obtained by concatenating the horizontal planes (cf. (4.11)) in a series, and the matrix form of $D \in \Re^{n \times lm}$ is obtained by

connecting the lateral planes (cf. (4.12)) of the array in a row.

The difficulty with randomization tests is often with the test statistics. In order to explain any existing structure of the interactions between three factors, the residual array D is decomposed using the three-way principal component analysis. In the decomposition, although $l+m+n$ components are required to reproduce the total system of the interaction, often much of this structure can be accounted for by a smaller number, $s+t+u$, of the principal components. If so, there is almost as much information in the $s+t+u$ components as there is in the original as $l+m+n$ variables. The $s+t+u$ principal components can then replace the initial $l+m+n$ variables, and the original residuals. These condensed components of the residuals often shed light on relationships that were not previously suspected and thereby allow interpretations that would not ordinarily result.

Since the core array reflects the interactions between principal components of the factors, all interactions between the three factors should intuitively be summarized by some functions of core elements c_{pqr}. In particular, from the relation in (3.11a), the core array represents a partitioning of the overall fitted sum of squares of residuals into small units which can be attributed to each component. Hence, the proportion of the squared core elements, with respect to residual sums of squares, seems to be the natural choice as a test statistic for testing the null hypothesis of no three-factor interaction. The reasonable test statistic, say, for the case of $(s+t+u)$ terms, is the test of

$$H_0:\ c_{111}^2 = \cdots = c_{stu}^2 = 0,\ \text{for } p=1,\cdots s,\ q=1,\cdots t,\ r=1,\cdots u, \tag{3.13}$$

using

$$\frac{\sum_{p=1}^{s}\sum_{q=1}^{t}\sum_{r=1}^{u}\hat{c}_{pqr}^2}{\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n}d_{ijk}^2} = \frac{\sum_{p=1}^{s}\sum_{q=1}^{t}\sum_{r=1}^{u}\hat{c}_{pqr}^2}{||D||^2} = \frac{\sum_{p=1}^{s}\sum_{q=1}^{t}\sum_{r=1}^{u}\hat{c}_{pqr}^2}{SS(ABC)}. \tag{3.14}$$

The ratio measures the sum of squares accounted for by the $s+t+u$ core terms with respect to the residual sum of squares. As stated earlier, part of the interaction may well be random error. Consequently, in practice, the partitioning of the interaction into

multiplicative terms is carried out partially so only a few multiplicative terms of the $\eta_{ijk} = c_{pqr} g_{ip} h_{jq} e_{kr}$ type are retained. Generally one or two such terms would summarize the interaction structure properly.

In this section, randomization tests will be applied for testing a single multiplicative term rather than testing the group of the components simultaneously. Thus if the result is significant, only the largest multiplicative term would be retained in the model (3.5) and the remaining terms are pooled together and considered as experimental error. The extension to more than a single term would involve a stepwise test procedure with more computation.

Let c^2 or $c^2_{(1)}$ be the largest squared element of core array C of the residual (3.6),

$$c^2 = c^2_{(1)} = \max \{c^2_{pqr} \mid 1 \le p \le s,\ 1 \le q \le r,\ 1 \le r \le u\}. \tag{3.15}$$

For the case when c^2 dominates all other c^2_{pqr}, i.e. $c^2/||C||^2$ is near 1 or much greater than the ratios for the other core elements, then the model (3.5) can be simplified as

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + c\, g_i h_j e_k + \epsilon_{ijk} \tag{3.16}$$

where the parameters are defined as in the model (3.1) with

$$\epsilon_{ijk} \sim \mathrm{NID}(0, \sigma^2)$$

$$\sum_i \alpha_i = \sum_j \beta_j = \sum_k \gamma_k = \sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = \cdots = \sum_k (\beta\gamma)_{jk} = 0.$$

The multiplicative term $c\, g_i h_j e_k$ represents a three-factor interaction term satisfying

$$\sum_i g_i = \sum_j h_j = \sum_k e_k = 0 \quad \text{and} \quad \sum_i g_i^2 = \sum_j h_j^2 = \sum_k e_k^2 = 1 \tag{3.17}$$

where $i = 1, \cdots, l$, $j = 1, \cdots, m$, and $k = 1, \cdots, n$. For this model, the hypothesis for testing three-factor interaction is

$$H_0: \; c^2 = 0 \quad \text{versus} \quad H_1: \; c^2 > 0 . \tag{3.18}$$

For the testing this hypothesis, the following two test statistics are considered first:

$$S_1 = \frac{\hat{c}^2}{\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}^2} = \frac{\hat{c}^2}{||D||^2} \tag{3.19}$$

and

$$S_2 = \frac{\hat{c}_{111}^2}{\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}^2} = \frac{\hat{c}_{111}^2}{||D||^2} \tag{3.20}$$

where $\hat{c}^2 = \max\{ \hat{c}_{pqr}^2 \mid 1 \le p \le s, \; 1 \le q \le r, \; 1 \le r \le u\}$, i.e., the largest estimated squared core element. Notice that the ranges of the statistics S_1 and S_2 are the interval [0, 1] since these are proportions.

To test the null hypothesis, (3.18), of no interaction, the statistic S_1 is the most natural choice for a test statistic since it measures the proportion of the largest core element. However, in the beginning of the research, S_2 was used as a test statistic since c_{111} represents the interaction between the first principal components of the three factors and so the magnitude of c_{111}^2 is expected to be the largest. As a result, this should well summarize the interaction structure in the residuals, if any. It turns out, in general, from the results based on some of the simulated data, that c_{111}^2 is not necessarily always the largest. This leads to a bimodality problem as discussed below.

To overcome this, in addition to the statistics S_2, which is based on the core elements, the statistics based on the eigenvalues, L(3.27), M(3.28), and N(3.29), are also studied. As an analogue of Section 2.2.3, we have

$$||D||^2 = \sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}^2 = \sum_{p=1}^{l-1} \hat{\lambda}_p = \sum_{q=1}^{m-1} \hat{\mu}_q = \sum_{r=1}^{n-1} \hat{\nu}_r \tag{3.21}$$

where

$\hat{\lambda}_1 > \cdots > \hat{\lambda}_{l-1}$ are the eigenvalues of $P(H,E) = D(HH' \otimes EE')D'$, where $D \in \Re^{l \times nm}$ (3.22)

$\hat{\mu}_1 > \cdots > \hat{\mu}_{m-1}$ are the eigenvalues of $Q(G,E) = D(EE' \otimes GG')D'$, where $D \in \Re^{m \times ln}$ (3.23)

$\hat{\nu}_1 > \cdots > \hat{\nu}_{n-1}$ are the eigenvalues of $R(G,H) = D(GG' \otimes HH')D'$, where $D \in \Re^{n \times lm}$. (3.24)

By combining the equations in (3.11) and (3.21),

$$\sum_{p=1}^{s}\sum_{q=1}^{t}\sum_{r=1}^{u}\hat{c}^2_{pqr} = \sum_{p=1}^{s}\hat{\lambda}_p = \sum_{q=1}^{t}\hat{\mu}_q = \sum_{r=1}^{u}\hat{\nu}_r. \tag{3.25}$$

Furthermore, we obtain the following relations between the the largest eigenvalue of the residual matrices and the core elements:

$$\begin{aligned}
\hat{\lambda}_1 &= \sum_{q=1}^{t}\sum_{r=1}^{u}\hat{c}^2_{1qr}, \\
\hat{\mu}_1 &= \sum_{p=1}^{s}\sum_{r=1}^{u}\hat{c}^2_{p1r}, \\
\hat{\nu}_1 &= \sum_{p=1}^{s}\sum_{q=1}^{t}\hat{c}^2_{pq1},
\end{aligned} \tag{3.26}$$

where, $\hat{\lambda}_1$, $\hat{\mu}_1$, and $\hat{\nu}_1$ are the largest eigenvalues of the corresponding cross product matrices D in (3.22)-(3.24) and $s = t = u$ are often one or two. Notice that all three of the summands in the equations (3.26) contain c^2_{111}. As a result, when c^2_{111} is much larger than the other squared core elements, then c^2_{111} is approximately equal to any of the eigenvalues: $\hat{\lambda}_1$, $\hat{\mu}_1$, and $\hat{\nu}_1$. In this case, these eigenvalues are about the same as well. Thus, S_2 can be approximated by any of

$$L = \frac{\hat{\lambda}_1}{||D||^2} = \frac{\hat{\lambda}_1}{\sum_{p=1}^{l-1} \hat{\lambda}_p}, \tag{3.27}$$

$$M = \frac{\hat{\mu}_1}{||D||^2} = \frac{\hat{\mu}_1}{\sum_{q=1}^{m-1} \hat{\mu}_q}, \text{ and} \tag{3.28}$$

$$N = \frac{\hat{\nu}_1}{||D||^2} = \frac{\hat{\nu}_1}{\sum_{r=1}^{n-1} \hat{\nu}_r}. \tag{3.29}$$

In particular, the best approximation in this case would be

$$S_2 \simeq \min \{ L, M, N \} \tag{3.30}$$

since c^2_{111} will be closest to the minimum of the eigenvalues in (3.26), i.e., $c^2_{111} \simeq$ min ($\hat{\lambda}_1$, $\hat{\mu}_1$, and $\hat{\nu}_1$). The further justification for these being suitable statistics is given in Section 3.4.2.

To compare the statistics, the results of the randomization test following the steps in Section 3.3.2 are presented now. First, the results of the test for the test statistic S_2 are summarized below:

(1) When three-factor interaction is present in the data, the observed values s_0 of S_2 is large. As expected, increasing the interaction effects leads to larger proportions accounted for by c^2_{111}. In this case, $\hat{c}^2_{111} = \hat{c}^2_{(1)}$ holds.

(2) On the other hand, for the data with no three factor-interaction, the value s_0 of S_2 is quite small. Decreasing the interaction, leads to smaller values of S_2. For this case, $\hat{c}^2_{111} = \hat{c}^2_{(1)}$ still holds most of the time, but not always. This leads to a bimodality problem.

(3) The randomization distribution of the test statistic S_2 reveals bimodality. The histograms of the test statistic show two modes; the first mode is located near the

median (often around .18 to .24) and the second mode is, near zero, on the far end of the left tail of the distribution (see Figure 3). Most sample values s_0 are clustered near the first mode of the randomization distribution. About 10-15% of time, the values s_0 of S_2 are gathered around the second peak of the distribution. For these values, as expected, not only $\hat{c}^2_{111} \neq \hat{c}^2_{(1)}$, but also their magnitudes are rather small.

(4) The significance levels or the p-values of s_0 are reasonable. The range of p-values for the simulated data with no interaction is between .288 to .604 and for the data with interaction the range of p-values is between .004 to .002 using 500 iterations.

The other statistic for testing interaction is S_1, the proportion of the largest squared core element with respect to the residual sum of squares. The following are the results of the simulation for the randomization test for S_1:

(1) The results of (1) for the statistic S_2 also hold for S_1.

(2) For the data with no three-factor interaction, the value s_0 of S_1 is small. Again, decreasing the interaction effects leads to smaller values of S_1.

(3) The randomization distribution of S_1 is unimodal. The histogram of the test statistic value s_0 of S_1, shows a bell shaped curve with mild skewness to the right. Since S_1 always picks up the proportion of the maximum of the squared core elements, the bimodality problem is gone.

(4) The p-value of S_1 is slightly smaller but about the same as that for S_2.

(5) S_1 is the likelihood ratio test statistic, as discussed in section 3.5, for the hypothesis,

$$H_0: \; c^2 = c^2_{(1)} = 0 \quad \text{versus} \quad H_1: \; c^2 = c^2_{(1)} > 0 .$$

Now we briefly note the following results for the statistics for L, M, and N:

(1) The outcomes (1) and (2) of the results for S_1 still hold for these statistics.

(2) Most notably, the randomization distributions, as shown in Figures 4 to 6, display much smoother and more symmetric bell shaped curves as compared to those of S_1 and S_2.

(3) The p-values for the data with large interaction are consistent with those of the statistics S_1 and S_2. However, for the data under the null hypothesis, the p-values for these statistics are more significant than those of S_1 and S_2: for example, the p-values under null hypothesis based on these statistics are between .01 to .09 and the p-value of S_1 and S_2 are near .30 (cf. Tables 7 and 8).

From the above observations, the statistics S_1, L, M, and N appear to be appropriate statistics for the randomization test. Intrinsically, however, the statistic S_1 seems to be the most suitable choice of the test statistic for the randomization test since it tests the proportion of the largest core element which directly relates to the interaction structure. Notice that, except for the bimodality of the distribution of S_2, both of the test statistics, S_1 and S_2 are basically giving the same results as far as the testing is concerned. In Section 3.4, an approximate test based on the test statistic S_2 will be presented under the restriction $c^2_{(1)} = c^2_{111}$.

3.3.4 Examples of randomization test

The randomization test procedure is demonstrated using the test statistic S_1. Two artificial examples are considered to illustrate the procedure.

(A) Example 1

First, we look at an example with simulated data which is generated under the alternative hypothesis H_1: $c^2 > 0$. The model for Example 1 is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijk},$$

where $\mu = 25$, $\underline{\alpha}' = (4,\ 3,\ 2,\ 0,\ -4,\ -5)$, $\underline{\beta} = (3,\ 2,\ 1,\ -2,\ -4)$, $\underline{\gamma} = (2,\ 1,\ 0,\ -3)$, with $(\alpha\beta)_{ij} = \alpha_i\beta_j$, and $(\alpha\beta\gamma)_{ijk} = .3\alpha_i\beta_j\gamma_k$. Table 1 shows the $6 \times 5 \times 4$ data. Following the steps in Section 3.3.2, the procedure and the results of a randomization test for the data are as follows:

1. For the given dataset, the standard ANOVA model is fitted and the result is shown in Table 2. From this fit, the residual matrices in Table 3 are obtained. Notice that the residual sum of squares, SSE, accounts about 20% of the total variation.

2. Using a SAS/IML program, the observed test statistic s_0 of S_1 is computed from the obtained residual array D in step (1): $s_0 = .98$.

3. The residuals are randomly allocated and the statistic value s_1 is computed for each permutation. If $s_1 \geq s_0$, then the counter is increased.

4. Step (3) is repeated a large number of times e.g., 500 to find a sample of values from the distribution of D that occurs by randomly allocating the scores actually observed to 120 residuals. These estimates yield the randomization distribution shown in Figure 2. The distribution looks fairly normal but with mild skewness to the right. Since $s_0 = .98$ is unusually large from the randomization distribution, the data are unlikely to have arisen if the null model is true and one can conclude that the alternative hypothesis is more plausible.

5. To summarize the randomization results, the p-value, the proportion of s_1 values that are as large as s_0 with respect to the number of permutations is calculated to be $1/500 = .002$.

6. Conclusion: We reject the null hypothesis, H_0: $c^2 = 0$, and support the alternative hypothesis, H_1: $c^2 > 0$. Thus the randomization test indicates that there is a significant three-factor interaction which can be best summarized by a single multiplicative term and the model (3.16) is suggested as a suitable model for the data.

Table 1. Dataset for Example 1 under alternative model

c1	b1	b2	b3	b4	b5
a1	51.95	45.60	39.08	17.71	-0.67
a2	48.49	41.04	35.31	17.76	4.72
a3	42.23	38.03	32.63	21.48	10.25
a4	30.17	28.16	29.43	24.85	24.87
a5	8.20	11.58	16.57	32.86	46.29
a6	0.80	8.09	15.26	36.60	48.67

c2	b1	b2	b3	b4	b5
a1	49.32	43.48	35.30	17.97	5.84
a2	45.08	39.53	33.99	19.70	10.74
a3	36.12	35.38	31.67	21.74	12.77
a4	30.52	28.16	25.92	24.72	22.40
a5	9.25	14.19	18.91	30.52	38.58
a6	3.35	10.35	15.06	30.38	42.25

c3	b1	b2	b3	b4	b5
a1	44.31	38.56	36.35	18.91	9.69
a2	38.99	38.29	31.89	20.90	14.54
a3	35.49	34.70	30.47	22.43	14.93
a4	28.68	26.72	25.30	23.58	21.32
a5	10.64	14.42	16.67	27.49	32.45
a6	7.48	13.33	16.84	28.27	36.58

c4	b1	b2	b3	b4	b5
a1	30.36	30.18	25.87	22.57	20.52
a2	30.44	29.01	28.34	22.73	18.86
a3	25.79	26.54	25.55	21.63	18.99
a4	22.59	23.82	23.61	18.28	19.65
a5	18.74	18.60	20.92	16.94	16.71
a6	18.31	17.89	17.38	18.48	14.48

Table 2. Analysis of variance table for Example 1

Source	DF	Type III SS	Mean Square	F Value	Pr > F
A	5	1482.2198	296.4440	6.01	0.0001
B	4	793.5262	198.3816	4.02	0.0059
C	3	398.4869	132.8290	2.69	0.0540
A*B	20	9720.3917	486.0196	9.86	0.0001
A*C	15	20.7295	1.3820	0.03	1.0000
B*C	12	11.1497	0.9291	0.02	1.0000
Error	60	2958.284	49.305		
Corrected Total	119	15384.788			

R-Square	C.V.	Root MSE	Y Mean
0.807714	27.94495	7.0217	25.127

Table 3. Residual matrices for Example 1

c1	b1	b2	b3	b4	b5
a1	5.68	4.93	3.28	-3.30	-10.59
a2	6.10	3.50	1.92	-3.59	-7.93
a3	4.65	2.76	0.51	-2.46	-5.46
a4	-0.87	-0.54	0.95	-0.50	0.96
a5	-6.28	-4.82	-3.83	3.70	11.22
a6	-9.27	-5.84	-2.82	6.14	11.80

c2	b1	b2	b3	b4	b5
a1	3.94	2.57	0.27	-2.48	-4.30
a2	2.89	1.06	0.67	-1.78	-2.83
a3	0.46	0.91	1.34	-0.60	-2.12
a4	1.17	0.02	-1.00	0.74	-0.92
a5	-3.88	-1.99	-0.27	2.38	3.75
a6	-4.58	-2.57	-1.01	1.74	6.41

c3	b1	b2	b3	b4	b5
a1	0.33	-1.45	1.92	-0.91	0.11
a2	-1.75	0.78	-0.75	0.11	1.61
a3	0.34	0.24	-0.13	-0.17	-0.28
a4	1.13	-0.12	-0.61	0.63	-1.04
a5	0.05	0.28	-0.76	1.12	-0.68
a6	-0.09	0.27	0.34	-0.78	0.27

c4	b1	b2	b3	b4	b5
a1	-9.95	-6.05	-5.47	6.69	14.78
a2	-7.24	-5.34	-1.83	5.26	9.14
a3	-5.45	-3.91	-1.72	3.22	7.85
a4	-1.42	0.63	0.66	-0.86	0.99
a5	10.11	6.53	4.86	-7.20	-14.29
a6	13.94	8.14	3.49	-7.10	-18.47

Variable $= S_1$

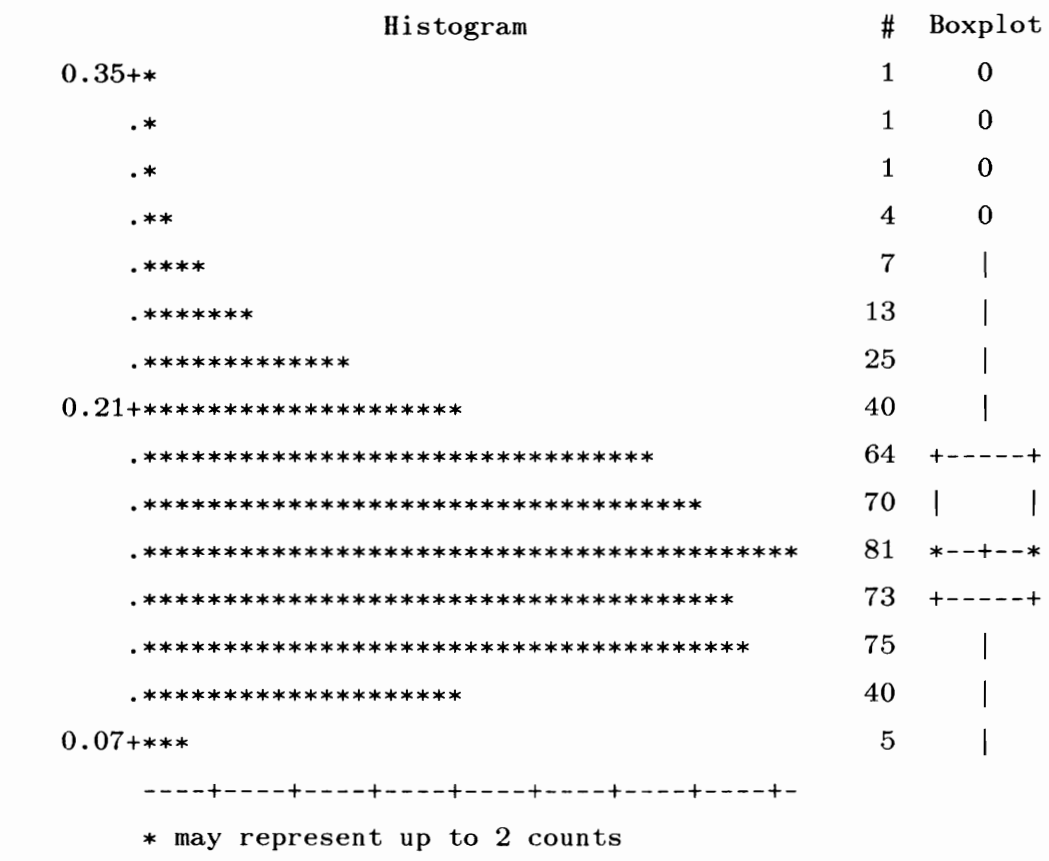


Figure 2. The distribution of test statistic S_1

(B) Example 2

Now we examine a data set which was simulated under the null hypothesis H_0: $c^2 = 0$. The dataset for Example 2 is based on the three-factor additive model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijk},$$

using the same main effect parameters as for Example 1. Table 4 shows the dataset. The procedure for a randomization test is the same as for Example 1. The results are summarized as follows:

1. For the dataset for Example 2, the ANOVA table shown in Table 5. From this fit, the residual matrices in Table 6 are obtained. Notice that SSE for this data accounts for less than 4% of the total variation. Again, the residuals are computed from this standard fit of analysis of variance.

2. Using the program, the observed test statistic s_0 of S_1 is computed and its value is $s_0 = .18147$.

3. The residuals are randomly permuted 500 times, the statistic s_1 is computed for each permutation and the number of cases for which this value is greater than or equal to s_0 is recorded.

4. Since $s_0 = .18$ looks like a typical value from the randomization distribution, we conclude that the allocation of the residuals to the 120 treatments combinations that occurred seems to be random.

5. The p-value for this data set is $148/500 = .296$.

6. For the dataset, we do not reject the null hypothesis. Thus the randomization test indicates that there is no significant three-factor interaction and we conclude model (3.2), which has no multiplicative term, is the proper model for dataset in Example 2.

Table 4. Dataset for Example 2 under null model

c1	b1	b2	b3	b4	b5
a1	32.7508	32.7989	32.6791	30.5122	24.9345
a2	34.0902	31.4398	30.5071	27.3603	23.9179
a3	32.6334	31.6266	29.4284	27.8793	23.0490
a4	30.1735	28.1645	29.4320	24.8505	24.8710
a5	27.4045	24.3796	22.9750	20.0627	20.6937
a6	24.7974	24.0929	23.2648	20.5989	16.6697

c2	b1	b2	b3	b4	b5
a1	33.7222	33.0768	30.1026	28.3681	26.6431
a2	33.3833	31.7321	30.0886	27.5042	26.3352
a3	28.3192	30.1826	29.0674	26.9441	23.1745
a4	30.5188	28.1624	25.9177	24.7228	22.4025
a5	24.8473	24.5859	24.1098	20.1177	17.7832
a6	22.8451	23.3480	21.5643	17.3850	16.2462

c3	b1	b2	b3	b4	b5
a1	32.3126	30.5642	32.3451	26.9125	25.6906
a2	29.9894	32.2866	28.8935	26.9035	26.5414
a3	29.4866	30.7020	28.4674	26.4305	22.9325
a4	28.6836	26.7239	25.3016	23.5829	21.3198
a5	22.6439	22.4210	20.6687	19.4914	16.4461
a6	22.4797	23.3313	21.8369	18.2718	16.5807

c4	b1	b2	b3	b4	b5
a1	29.1639	29.3787	25.4703	23.3691	22.1217
a2	29.5421	28.4075	28.0381	23.3288	20.0591
a3	25.1879	26.1359	25.3540	22.0329	19.7897
a4	22.5877	23.8152	23.6084	18.2763	19.6466
a5	19.9393	19.4016	21.3187	16.1405	15.1149
a6	19.8090	18.8913	17.8757	17.4801	12.4764

Table 5. Analysis of variance table of Example 2

Source	DF	Type III SS	Mean Square	F Value	Pr > F
A	5	1482.2198	296.4440	230.71	0.0001
B	4	793.5262	198.3816	154.39	0.0001
C	3	398.4869	132.8290	103.37	0.0001
A*B	20	20.2658	1.0133	0.79	0.7162
A*C	15	20.7295	1.3820	1.08	0.3974
B*C	12	11.1497	0.9291	0.72	0.7233
Error	60	77.0957	1.2849		
Total	119	2803.4736			

R-Square	C.V.	Root MSE	Y Mean
0.972500	4.511268	1.1335	25.127

Table 6. Residual matrices for Example 2

c1	b1	b2	b3	b4	b5
a1	-1.5219	0.1309	0.8802	1.5013	-0.9906
a2	0.6973	-0.0964	0.1193	0.0093	-0.7295
a3	1.0484	0.3586	-0.6934	-0.0557	-0.6579
a4	-0.8737	-0.5363	0.9465	-0.4990	0.9624
a5	0.9239	-0.0173	-1.4292	-1.0973	1.6200
a6	-0.2739	0.1605	0.1765	0.1413	-0.2044

c2	b1	b2	b3	b4	b5
a1	0.3431	0.1692	-0.9323	-0.0811	0.5011
a2	0.1858	-0.7419	-0.2334	0.0167	0.7728
a3	-1.3389	-0.2918	0.7428	0.6040	0.2840
a4	1.1661	0.0229	-1.0029	0.7358	-0.9221
a5	-0.2780	0.4111	0.9313	-0.0189	-1.0455
a6	-0.0781	0.4304	0.4945	-1.2565	0.4097

c3	b1	b2	b3	b4	b5
a1	0.3258	-1.4480	1.9186	-0.9096	0.1132
a2	-1.7476	0.7763	-0.7518	0.1114	1.6117
a3	0.3370	0.2392	-0.1326	-0.1663	-0.2774
a4	1.1282	-0.1153	-0.6056	0.6280	-1.0353
a5	0.0479	0.2786	-0.7643	1.1189	-0.6811
a6	-0.0912	0.2692	0.3357	-0.7824	0.2688

c4	b1	b2	b3	b4	b5
a1	0.8530	1.1479	-1.8666	-0.5106	0.3763
a2	0.8645	0.0621	0.8659	-0.1374	-1.6550
a3	-0.0464	-0.3060	0.0831	-0.3820	0.6513
a4	-1.4206	0.6286	0.6620	-0.8649	0.9949
a5	-0.6938	-0.6724	1.2622	-0.0027	0.1067
a6	0.4433	-0.8601	-1.0067	1.8976	-0.4741

Variable $= S_2$

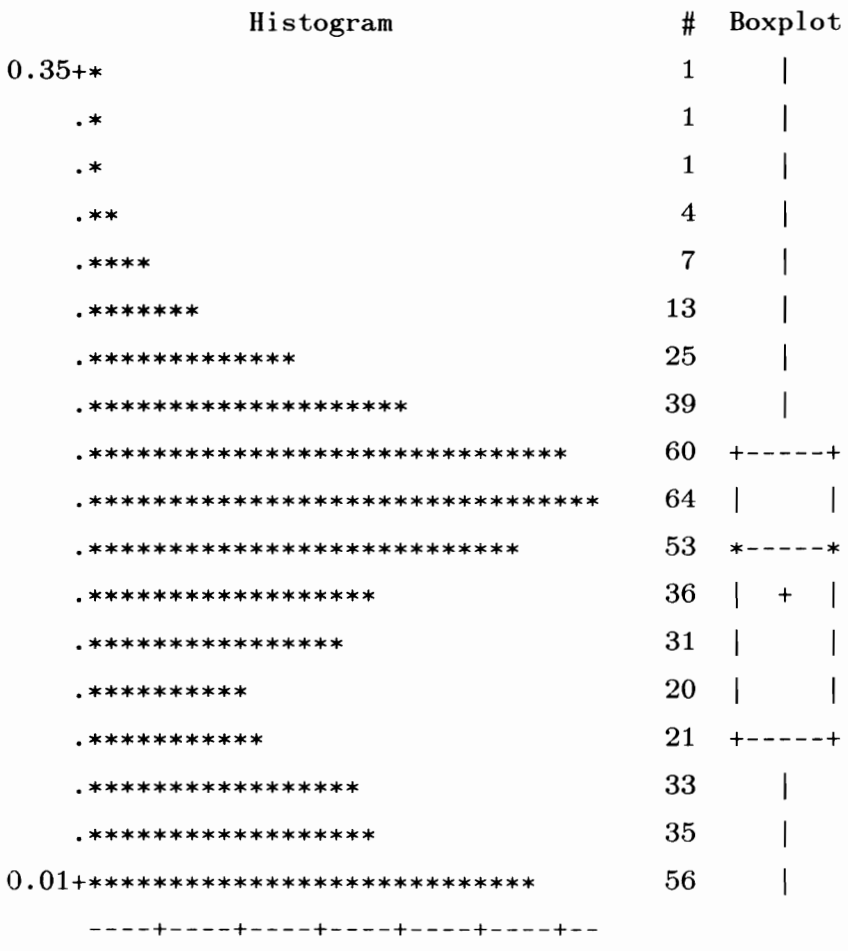


Figure 3. The distribution of the test statistic S_2

Variable $= L$

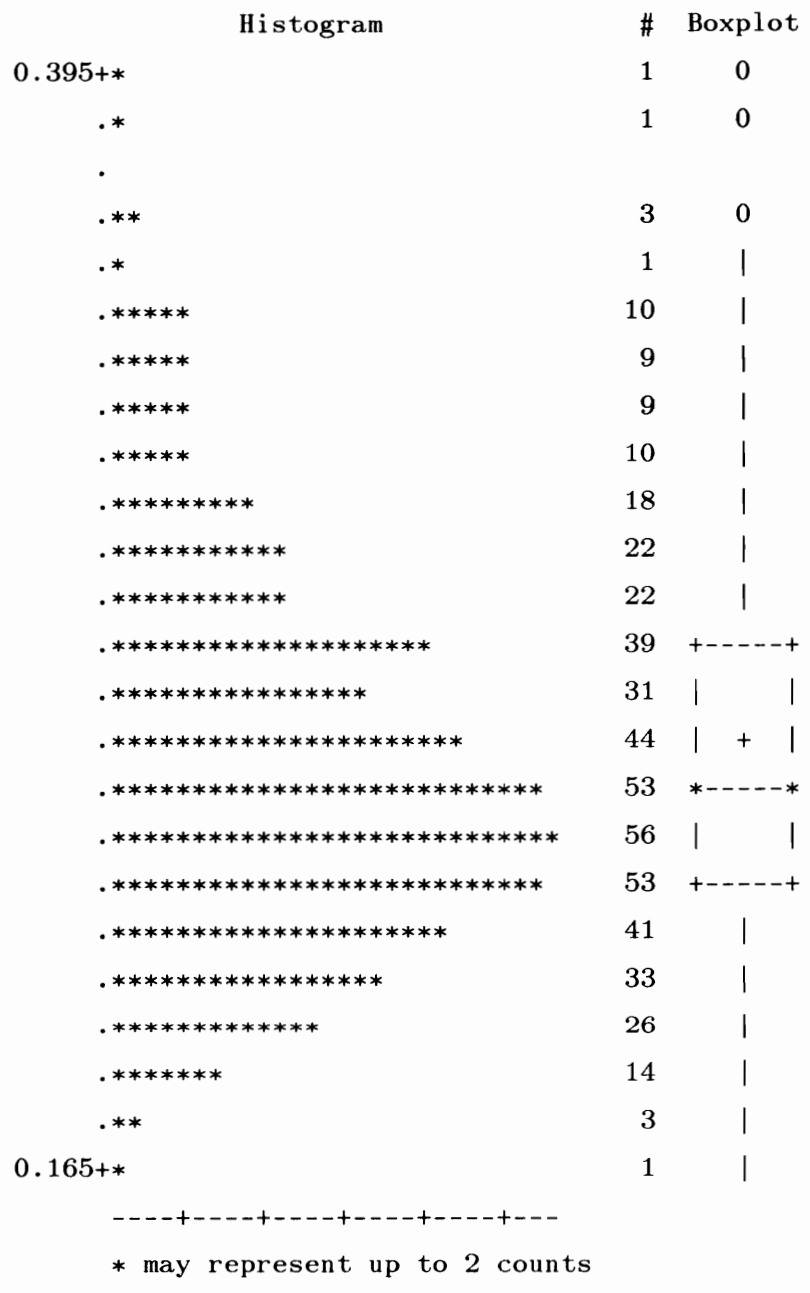


Figure 4. The distribution of the test statistic L

Variable = M

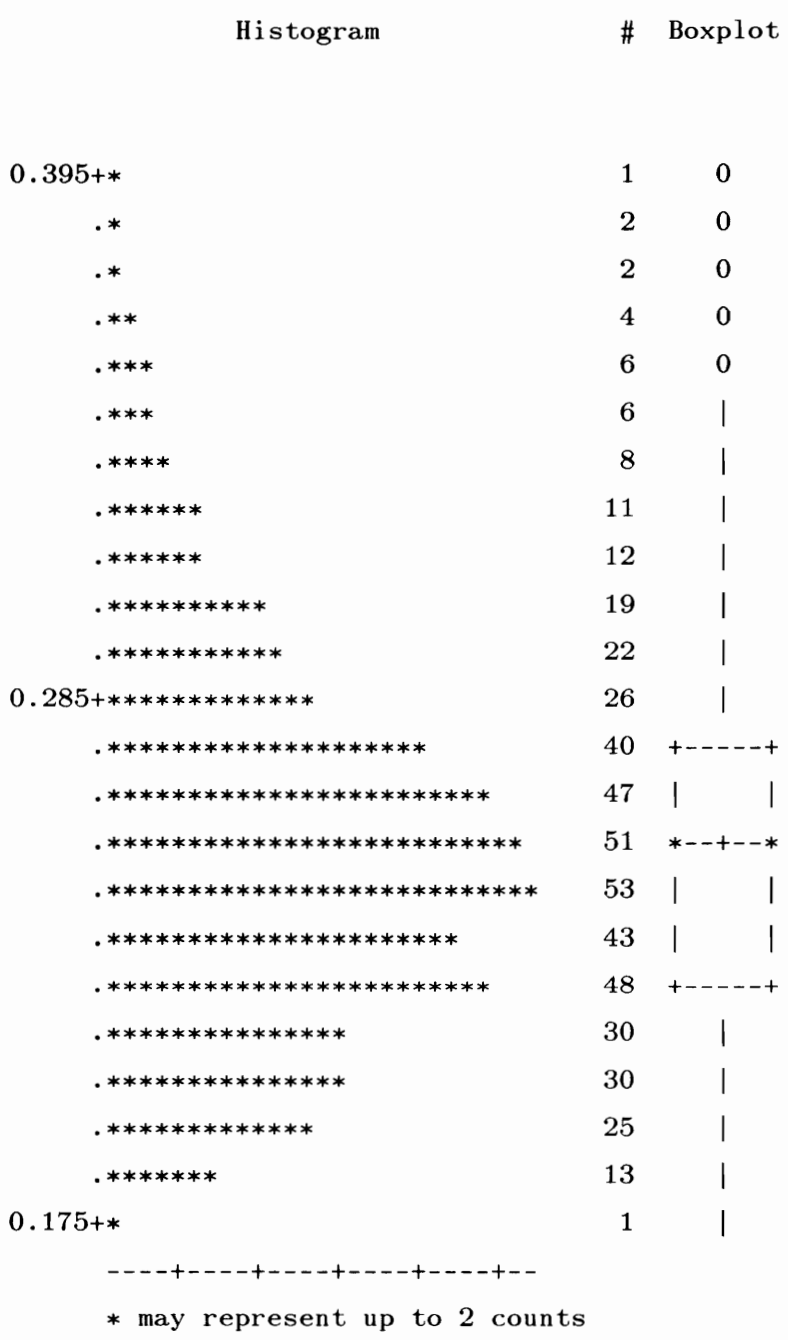


Figure 5. The distribution of the test statistic M

Variable = N

```
                Histogram                    #  Boxplot
  Stem Leaf                                  #  Boxplot
    42 018                                   3     0
    41 189                                   3     0
    40 17                                    2     |
    39 8                                     1     |
    38 222568                                6     |
    37 12335589                              8     |
    36 144679                                6     |
    35 01345899                              8     |
    34 000335666778899                      15     |
    33 0012334556667889                     16     |
    32 00111222233445566777778889           26     |
    31 000123455677899999999                21     |
    30 00111222333344555556667888899999     32  +-----+
    29 000000011112222334445556788899999    32  |     |
    28 00000111222223333344445555555567777889  38  |     |
    27 000000001122333333444455555566667777778888899999  47  *--+--*
    26 00000111123334444456666777778888888  36  |     |
    25 000011111122233444445555666667778888999999  42  |     |
    24 00000222333444444566677777889 99      31  |     |
    23 01111122223334444455555666666777779999 99  41  +-----+
    22 00111222344444445566667777888889 99   34     |
    21 11122223346677888899999              23     |
    20 0123344456666789                     16     |
    19 123333579                             9     |
    18 26                                    2     |
    17 99                                    2     |
       ----+----+----+----+----+----+----+----+----+--
  Multiply Stem.Leaf by 10**-2
```

Figure 6. The distribution of the test statistic N

Table 7. Summary table for randomization tests for Example 1

	Computed test statistic value	Randomization distribution	P-value
S_1	.978	Unimodal Slightly skewed	.002
S_2	.978	Bimodal	.002
L	.980	Unimodal Bell shaped	.002
M	.980	Unimodal Bell shaped	.002
N	.980	Unimodal Bell shaped	.002

Table 8. Summary table for randomization tests for Example 2

	Computed test statistic value	Randomization distribution	P-value
S_1	.181	Unimodal Slightly skewed	.296
S_2	.181	Bimodal	.288
L	.041	Unimodal Bell shaped	.010
M	.049	Unimodal Bell shaped	.044
N	.037	Unimodal Bell shaped	.092

For a complete illustration, the distribution of the test statistics S_2, L, M, and N are presented in Figures 3 to 6. The outcomes from the randomization tests along with the above results for S_1 are summarized in Tables 7 and 8.

Remarks on Randomization tests

The above randomization test is based on a single multiplicative term. In practice it is possible to have complicated interaction. This leads to a sequential approach to testing. If the null hypothesis of a single multiplicative term (3.18) is rejected, one must test for a second multiplicative component or one may be interested in testing for a group of components simultaneously to see if any other term(s) should belong(s) in the model. Concerning this issue, it is worth while to mention the following:

Case 1: If the p-value is significantly small, say less than .05, (i.e. c^2 dominates the core) and the proportion $||C||^2/||D||^2$ is large, e.g., greater than 90%, then the three-factor interaction in a data can be summarized well with one multiplicative term $c g_i h_j e_k$.

Case 2: If the p-value is small but not significantly small, yet the proportion $||C||^2/||D||^2$ is large, then compare the quantity of the test statistics

$$s_{(1)} = \frac{c^2_{(1)}}{||D||^2} \quad \text{and} \quad s_{(2)} = \frac{c^2_{(1)} + c^2_{(2)}}{||D||^2},$$

where $s_{(2)}$ is the proportion of the sum of the first two largest core element.

(i) If the difference between $s_{(1)}$ and $s_{(2)}$ is negligible, then no multiplicative term belongs in the model and the null model in (3.2) is supported.

(ii) If $s_{(2)}$ is much larger than $s_{(1)}$, then investigate the second multiplicative term for possible interaction structure.

Case 3: If the proportion $||C||^2/||D||^2$ is rather small, then the reduction on residual arrays are done crudely. In this case, the fit of the three-mode principal component analysis on residuals needs to be improved before the randomization test is performed. It may be possible to improve the fit by adding more components.

3.4. LIKELIHOOD RATIO TESTS

This section discusses likelihood ratio tests for testing the hypothesis of no three-factor interaction in a three-way classification model with a single observation per treatment combination. In particular, a likelihood ratio test statistic is derived to test for a single multiplicative term. In Section 3.4.1, the maximum likelihood estimates (MLE) are computed. Section 3.4.2 derives a likelihood ratio test statistic for testing the null hypothesis H_0: $c^2 = 0$. It turns out that the likelihood ratio test statistic is S_1 (3.19). Moreover, under the constraint, $c^2 = c_{111}^2$, the test statistic S_2(3.20) is equivalent to S_1. In Section 3.4.3, the distributional properties of an approximated likelihood ratio statistic is presented based on eigenvalues of the residuals to obtain percentage points. Notice that the analysis presented in this section is three-way analogue of Johnson and Graybill's (1972) results for the two-way table.

An alternative way of deriving the likelihood ratio test statistic is shown in Appendix B. In fact, the likelihood ratio statistic in (B.2) is for a general hypothesis (3.13) for testing groups of components.

For the model (3.16),

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + c\,g_i h_j e_k + \epsilon_{ijk}$$

a likelihood ratio test of the hypothesis H_0: $c^2 = 0$ versus H_1: $c^2 > 0$ will be derived below. If a result of the test shows that there is no three-factor interaction, then the usual analysis of main effects and two-factor interactions would be made. If a three-factor interaction is suspected, the procedure is recommended for

(i) obtaining a multiplicative term which is separated from the random error
(ii) acquiring a suitable estimate of the error variance σ^2
(iii) getting insight into the main structure of the interaction.

It is known, from Theorem 1 of Kroonenberg and De Leeuw (1980), that the alternating least squares (ALS) estimates of c_{pqr}, $\underline{g}_p$, $\underline{h}_q$, and $\underline{e}_r$ are, respectively, $\hat{c}_{pqr}$, $\hat{\underline{g}}_p$, $\hat{\underline{h}}_q$, and $\hat{\underline{e}}_r$, where

$\hat{\underline{g}}_p = [\hat{g}_{ip}]$ is an eigenvector of P in (3.22) corresponding to the eigenvalue λ_p, $p = 1, \cdots, s$,

$\hat{\underline{h}}_q = [\hat{h}_{jq}]$ is an eigenvector of Q in (3.23) corresponding to the eigenvalue μ_q, $q = 1, \cdots, t$,

$\hat{\underline{e}}_r = [\hat{e}_{kr}]$ is an eigenvector of R in (3.24) corresponding to the eigenvalue ν_r, $r = 1, \cdots, u$,

(3.31)

$\hat{c}_{pqr}$ = the (p, q, r)th element of the core matrix C for the residual array D (3.6)

$$= \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} d_{ijk}\, g_{ip}\, h_{jq}\, e_{kr}. \tag{3.32}$$

In the next section, the maximum likelihood estimators are found based on these alternating least squares estimators along with the usual least squares estimators. We assume that these estimators are the iterative solutions which minimize

$$||D - \hat{D}||^2 = \sum_{i=1}^{l} \sum_{j=1}^{m} \sum_{k=1}^{n} (d_{ijk}^2 - \hat{d}_{ijk}^2)^2,$$

where

$$D = [d_{ijk}], \text{ with } d_{ijk} = \sum_{p=1}^{l} \sum_{q=1}^{m} \sum_{r=1}^{n} c_{pqr}\, g_{ip}\, h_{jq}\, e_{kr},$$

and in terms of the sample estimates

$$\hat{D} = [\hat{d}_{ijk}], \text{ with } \hat{d}_{ijk} = \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} c_{pqr}\, g_{ip}\, h_{jq}\, e_{kr},$$

with $s < l$, $t < m$, and $u < n$ (cf. 3.8).

3.4.1 Maximum likelihood estimators

In this section we give the maximum likelihood estimates of the parameters for the single multiplicative model (3.16) and use them in the next section to determine the statistic for testing the hypothesis H_0: $c^2 = 0$ versus H_1: $c^2 > 0$.

First we note that when Y, D, C, G, H, and E are given as before, the maximum value of $(\sum_{p=1}^{l} \sum_{q=1}^{m} \sum_{r=1}^{n} d_{ijk}\, g_i h_j e_k)^2$ with respect to g_i, h_j, and e_k subject to the restrictions in (3.17) is $c^2 = \max\{\, c^2_{pqr} \mid 1 \le p \le s,\ 1 \le q \le r,\ 1 \le r \le u\}$ and is attained when $\underline{g}$, $\underline{h}$, and $\underline{e}$ are defined as above (3.31).

Theorem 3.4.1: For notational convenience, write the model (3.16) as

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \tau_{ij} + \rho_{ik} + \phi_{jk} + c\, g_i h_j e_k + \epsilon_{ijk} \tag{3.33}$$

with $\epsilon_{ijk} \sim \text{NID}(0, \sigma^2)$. The maximum likelihood estimators of parameters for the model are

$\hat{\mu} = \bar{y}_{...}$

$\hat{\alpha}_i = \bar{y}_{i..} - \bar{y}_{...}$

$\hat{\beta}_j = \bar{y}_{.j.} - \bar{y}_{...}$

$\hat{\gamma}_k = \bar{y}_{..k} - \bar{y}_{...}$

$$\hat{\tau}_{ij} = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}$$

$$\hat{\rho}_{ik} = \bar{y}_{i.k} - \bar{y}_{i..} - \bar{y}_{..k} + \bar{y}_{...}$$

$$\hat{\phi}_{jk} = \bar{y}_{.jk} - \bar{y}_{.j.} - \bar{y}_{..k} + \bar{y}_{...} \qquad (3.34)$$

$\hat{g}_i = i$th component of $\underline{\hat{g}}_{p_0}$ $(= [\hat{g}_{ip_0}])$, the normalized eigenvector of P in (3.22) corresponding to the eigenvalue λ_{p_0}, $p_0 = 1, \cdots, s$,

$\hat{h}_j = j$th component of $\underline{h}_{q_0}$ $(= [\hat{h}_{jq_0}])$, the normalized eigenvector of of Q in (3.23) corresponding to the eigenvalue μ_{q_0}, $q_0 = 1, \cdots, t$,

$\hat{e}_k = k$th component of $\underline{e}_{r_0}$ $(= [\hat{e}_{kr_0}])$, the normalized eigenvector of R in (3.24) corresponding to the eigenvalue ν_{r_0}, $r_0 = 1, \cdots, u$,

$\hat{c} = \hat{c}_{p_o q_o r_o}$, where $\hat{c}^2_{p_o q_o r_o} = \max\{ \hat{c}^2_{pqr} \mid 1 \le p \le s,\ 1 \le q \le r,\ 1 \le r \le u\}$ and

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d^2_{ijk} - \hat{c}^2}{lmn} = \frac{\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} \hat{c}^2_{pqr} - \hat{c}^2}{lmn}$$

where d_{ijk} is the residual in (i,j,k) th cell of the array D (3.6).

Proof: Let $\underline{\theta}' = [\ \mu,\ \underline{\alpha}',\ \underline{\beta}',\ \underline{\gamma}',\ \underline{\tau}',\ \underline{\rho}',\ \underline{\phi}',\ c,\ \underline{g}',\ \underline{h}',\ \underline{e}',\ \sigma^2]$

The likelihood function of $\underline{\theta}$ is

$$L(\underline{\theta}) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{lmn}{2}} e^{-\frac{1}{2\sigma^2}\left\{\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n}[y_{ijk} - (\mu + \alpha_i + \beta_j + \gamma_k + \tau_{ij} + \rho_{ik} + \phi_{jk} + c g_i h_j e_k)]^2\right\}} \qquad (3.35)$$

First, we partition the sum of squares in the exponent as follows:

$$\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n}\left[\, y_{ijk} - (\mu+\alpha_i+\beta_j+\gamma_k+\tau_{ij}+\rho_{ik}+\phi_{jk}+c g_i h_j e_k)\right]^2$$

$$= \sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n}\left[\, d_{ijk} + y_{ijk} - (y_{ijk} - \bar{y}_{ij.} - \bar{y}_{i.k} - \bar{y}_{.jk} + \bar{y}_{i..} + \bar{y}_{.j.} + \bar{y}_{..k} - \bar{y}_{...})\right.$$

$$\left. - (\mu+\alpha_i+\beta_j+\gamma_k+\tau_{ij}+\rho_{ik}+\phi_{jk}+c g_i h_j e_k) \,\right]^2$$

$$= \sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n}\left[\, d_{ijk} - (\mu-\bar{y}_{...}) - (\alpha_i - \bar{y}_{i..} + \bar{y}_{...}) - (\beta_j - \bar{y}_{.j.} + \bar{y}_{...}) - (\gamma_k - \bar{y}_{..k} + \bar{y}_{...})\right.$$

$$\left. - (\tau_{ij} - \bar{y}_{ij.} + \bar{y}_{i..} + \bar{y}_{.j.} - \bar{y}_{...}) - (\rho_{ik} - \bar{y}_{i.k} + \bar{y}_{i..} + \bar{y}_{..k} - \bar{y}_{...}) - (\phi_{jk} - \bar{y}_{.jk} + \bar{y}_{.j.} + \bar{y}_{..k} - \bar{y}_{...}) - c g_i h_j e_k\right]^2$$

$$= \sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}^2 + lmn\,(\mu-\bar{y}_{...})^2 + mn\sum_{i=1}^{l}(\alpha_i - \bar{y}_{i..} + \bar{y}_{...})^2 + ln\sum_{j=1}^{m}(\beta_j - \bar{y}_{.j.} + \bar{y}_{...})^2$$

$$+ lm\sum_{k=1}^{n}(\gamma_k - \bar{y}_{..k} + \bar{y}_{...})^2 + n\sum_{i=1}^{l}\sum_{j=1}^{m}(\tau_{ij} - \bar{y}_{ij.} + \bar{y}_{i..} + \bar{y}_{.j.} - \bar{y}_{...})^2$$

$$+ m\sum_{i=1}^{l}\sum_{k=1}^{n}(\rho_{ik} - \bar{y}_{i.k} + \bar{y}_{i..} + \bar{y}_{..k} - \bar{y}_{...})^2 + l\sum_{j=1}^{m}\sum_{k=1}^{n}(\phi_{jk} - \bar{y}_{.jk} + \bar{y}_{.j.} + \bar{y}_{..k} - \bar{y}_{...})^2$$

$$+ \Big(\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} c g_i h_j e_k\Big)^2 - 2c\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}\, g_i\, h_j\, e_k \,\Big]^2$$

Writing the last two terms as

$$c^2 - 2c\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}\, g_i\, h_j\, e_k$$

$$= \Big(\, c - \sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}\, g_i\, h_j\, e_k\Big)^2 - \Big(\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk} g_i h_j e_k\Big)^2,$$

the likelihood function becomes

$$L(\underline{\theta}) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{lmn}{2}} \exp\left\{ -\frac{1}{2\sigma^2}\left[\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}^2 + lmn\,(\mu-\bar{y}_{...})^2 + mn\sum_{i=1}^{l}(\alpha_i - \bar{y}_{i..}+\bar{y}_{...})^2 \right.\right.$$

$$+ ln\sum_{j=1}^{m}(\beta_j-\bar{y}_{.j.}+\bar{y}_{...})^2 + lm\sum_{k=1}^{n}(\gamma_k-\bar{y}_{..k}+\bar{y}_{...})^2 + n\sum_{i=1}^{l}\sum_{j=1}^{m}(\tau_{ij}-\bar{y}_{ij.}+\bar{y}_{i..}+\bar{y}_{.j.}-\bar{y}_{...})^2$$

$$+ m\sum_{i=1}^{l}\sum_{k=1}^{n}(\rho_{ik}-\bar{y}_{i.k}+\bar{y}_{i..}+\bar{y}_{..k}-\bar{y}_{...})^2 + l\sum_{j=1}^{m}\sum_{k=1}^{n}(\phi_{jk}-\bar{y}_{.jk}+\bar{y}_{.j.}+\bar{y}_{..k}-\bar{y}_{...})^2$$

$$\left.\left. + \left(c - \sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}\,g_i\,h_j\,e_k\right)^2 - \left(\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}g_ih_je_k\right)^2 \right] \right\}$$

$$\leq \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{lmn}{2}} \exp\left\{ -\frac{1}{2\sigma^2}\left[\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}^2 - \left(\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}\,g_i\,h_j\,e_k\right)^2 \right] \right\}$$

Now the last expression can be written as

$$y^t\,(1+x)^t\,e^{-(1+x)t}$$

by setting

$$y = \frac{1}{2\pi(1+x)\sigma^2}, \quad t = \frac{lmn}{2}, \text{ and}$$

$$1+x = \frac{1}{lmn\,\sigma^2}\left[\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}^2 - \left(\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}\,g_i\,h_j\,e_k\right)^2 \right]$$

Notice that for all $x > -1$,

$$(1+x)^t\,e^{-(1+x)t} \leq e^{-t}$$

since $(1+x)\,e^{-x} \leq e^{-1}$.

Thus the corresponding maximum value of the likelihood function is,

$$L(\underline{\theta}) \leq y^t\,(1+x)^t\,e^{-(1+x)t}$$

$$\leq (ye^{-1})^t$$

$$= \left(\frac{1}{2\pi(1+x)\sigma^2}\right)^{\frac{lmn}{2}} e^{-\frac{lmn}{2}}$$

$$= e^{-\frac{lmn}{2}} \left\{\frac{lmn}{2\pi\left[\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}^2 - \left(\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}\,g_i\,h_j\,e_k\right)^2\right]}\right\}^{\frac{lmn}{2}}.$$

Since the maximum value of $\left(\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}g_ih_je_k\right)^2$ with respect to g_i, h_j, and e_k is $\hat{c}^2$,

$$L(\underline{\theta}) \leq e^{-\frac{lmn}{2}} \left[\frac{lmn}{2\pi\left(\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}^2 - \hat{c}^2\right)}\right]^{\frac{lmn}{2}}. \tag{3.36}$$

To show that this maximum value is attained for the maximum likelihood estimators given in the theorem one needs only to substitute these estimators for the parameters in the likelihood equation. □

3.4.2 Likelihood ratio test statistic (LRTS)

In this section a likelihood ratio test statistic of the hypothesis H_0: $c^2 = 0$ versus H_1: $c^2 > 0$ is determined using the maximum likelihood estimates of the previous section. The likelihood ratio test for the hypothesis with more than a single term is discussed in Appendix C.

Let $L(\hat{\omega})$ be the maximum value of the likelihood function of the sample where the parameters $\underline{\theta}$ are restricted to values specified by H_0, and let $L(\hat{\Omega})$ be the maximum value of the likelihood function of the sample where the parameters may take on any value specified by the union of H_0 and H_1. Then the critical region consists of those sample outcomes for which

$$\Lambda^* = \frac{L(\hat{\omega})}{L(\hat{\Omega})} < k^{**}$$

where k^{**} is chosen so that P(type I error) = α.

The parameters under $H_0 \cup H_1$ are

$$\underline{\theta}_\Omega = [\ \mu,\ \underline{\alpha},\ \underline{\beta},\ \underline{\gamma},\ \underline{\tau},\ \underline{\rho},\ \underline{\phi},\ c,\ \underline{g},\ \underline{h},\ \underline{e},\ \sigma^2\ ],$$

and under H_0 are

$$\underline{\theta}_\omega = [\ \mu,\ \underline{\alpha},\ \underline{\beta},\ \underline{\gamma},\ \underline{\tau},\ \underline{\rho},\ \underline{\phi},\ \sigma^2\ ].$$

Thus under H_0, the usual least squares estimators (3.34) for a three-way fixed ANOVA model with no three-factor interaction term are the maximum likelihood estimators for the parameters $\underline{\theta}_\omega$ as the model (3.33) assumes normal errors. Under H_1, Theorem 3.4.1 gives the maximum likelihood estimators for the parameters $\underline{\theta}_\Omega$. Thus, the likelihood ratio test statistic of H_0: $c^2 = 0$ versus H_1: $c^2 > 0$ is

$$\Lambda^* = \frac{L(\hat{\omega})}{L(\hat{\Omega})} = \frac{e^{-\frac{lmn}{2}} \left[\frac{lmn}{2\pi \sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}^2} \right]^{\frac{lmn}{2}}}{e^{-\frac{lmn}{2}} \left[\frac{lmn}{2\pi \left(\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}^2 - \hat{c}^2 \right)} \right]^{\frac{lmn}{2}}}$$

$$= \left(\frac{\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}^2 - \hat{c}^2}{\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}^2} \right)^{\frac{lmn}{2}}.$$

So, we have the following theorem:

Theorem 3.4.2: For the model (3.33), the likelihood ratio test statistic of H_0: $c^2 = 0$ versus H_1: $c^2 > 0$ is

$$\Lambda^* = \left(\frac{\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}^2 - \hat{c}^2}{\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}^2} \right)^{\frac{lmn}{2}}. \tag{3.37}$$

Thus the test is: Reject H_0 in favor of H_1 if and only if

$$(\Lambda^*)^{\frac{2}{lmn}} = \frac{\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}^2 - \hat{c}^2}{\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}^2} < k^{**}$$

or equivalently, reject H_0 iff

$$\Lambda = \frac{\hat{c}^2}{\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}^2} = \frac{\hat{c}^2}{||D||^2} > k^* \tag{3.38}$$

where $\Lambda = 1-(\Lambda^*)^{\frac{2}{lmn}}$. Notice that the likelihood ratio test statistic Λ is the same as the randomization test statistic S_1 (3.19).

For a type I error probability of α, the hypothesis H_0: $c^2 = 0$ is rejected in favor of H_1: $c^2 > 0$, if $\Lambda > k^*$, where k^* is such that

$$P(\Lambda > k^* \mid H_0) = \alpha \quad \text{for a given } \alpha.$$

Thus, for example, from (3.21), $\Lambda > k$ is equivalent to

$$\Lambda = \frac{\hat{c}^2}{||D||^2} = \frac{\hat{c}^2}{\sum_{p=1}^{l-1} \hat{\lambda}_p} > k^*$$

where, k^* is such that $P(\Lambda > k^* \mid H_0) = \alpha$ for the given α.

To test the hypothesis based on the statistic, Λ, we need its distribution which is unknown. The likelihood ratio test statistic is based on the estimated squared core elements $\hat{c}^2_{pqr}$, $p = 1,\cdots, = s$, $q = 1,\cdots, = t$, $r = 1,\cdots, = u$, and as stated earlier, these are neither distributed independently nor as chi-square variables. One way to solve this problem is to construct tables of percentage points or critical values of Λ using Monte Carlo methods. An alternative way is to find a test statistic, which closely approximates Λ, and derive the distribution for the statistic.

First, we find a statistic which nearly approximates the likelihood ratio statistic Λ under the condition that $c^2 = c^2_{111}$ dominates the core, which is the often the case, where $c^2 = c^2_{(1)} = \max\{ c^2_{pqr} \mid 1 \le p \le s, 1 \le q \le r, 1 \le r \le u\}$ as defined in (3.15).

Thus,

$$\Lambda = \frac{\hat{c}^2}{||D||^2}$$

can be approximated by any of the statistics L in (3.27), M in (3.28), and N in (3.29). In particular, the best approximation in this case would be

$$\Lambda \simeq \min \{ L, M, N\} \tag{3.39}$$

since $c^2_{(1)} = c^2_{111}$ will be closest to the minimum of the eigenvalues in (3.39), i.e., $c^2 = c^2_{111} \simeq \min \{\hat{\lambda}_1, \hat{\mu}_1, \hat{\nu}_1\}$. Hence, without loss of generality, assuming that $\hat{\lambda}_1$ is the smallest of the three eigenvalues, we will study the distribution of L, instead of Λ, as the distribution of Λ is difficult to find.

As a result, we now test the hypothesis

$$H_0: \lambda_1 = 0 \quad \text{versus} \quad H_1: \lambda_1 > 0 \tag{3.40}$$

instead of H_0: $c^2 = 0$ under the restriction $c^2 = c^2_{111}$. Thus the test based on L is: reject H_0: $\lambda_1 = 0$ iff

$$L > k \tag{3.41}$$

where k is such that $P(L > k \mid H_0) = \alpha$ for a given α.

3.4.3 Approximate distribution of LRTS

In this section, the distribution of the statistic L will be derived as the approximate distribution of the likelihood ratio test statistic Λ. The investigation utilizes the distribution of the eigenvalues of the Wishart matrix as an extension of and analogue to the two-way case by Johnson and Graybill (1972) and Corsten and van Eijnsbergen (1972).

Here and in what follows, for the computational simplicity, we are using DD' and its

eigenvalues instead of the three-mode principal component analysis approximation P and its eigenvalues (3.22). The values of these are very close to each other as the PCA closely approximates the residuals. Also for notational convenience, we use the same notation $\hat{\lambda}_p$ for the eigenvalues. First we examine the joint distribution of $\hat{\lambda}_1, \cdots, \hat{\lambda}_{l-1}$ of the sum of squares and cross product matrix $DD' \in \Re^{l \times l}$. In fact, the joint distribution of $\hat{\lambda}_1, \cdots, \hat{\lambda}_{l-1}$ is the same as the joint distribution of the eigenvalues of a Wishart matrix.

Joint distribution of $\hat{\lambda}_1, \cdots, \hat{\lambda}_{l-1}$

Let $l \leq mn$, without loss of generality. From (3.6) and (3.12), D can be written as

$$\begin{aligned} D &= [\, d_{ij1} \mid d_{ij2} \mid \cdots \mid d_{ijn} \,] \\ &= (I_l - \tfrac{1}{l} \underline{1}_l \underline{1}_l')\, Y\, [\, (I_n - \tfrac{1}{n} \underline{1}_n \underline{1}_n') \otimes (I_m - \tfrac{1}{m} \underline{1}_m \underline{1}_m')] \\ &= H_l Y (H_n \otimes H_m), \end{aligned} \tag{3.42}$$

where $H_l = I_l - \frac{1}{l} \underline{1}_l \underline{1}_l'$, I_l is the $l \times l$ identity matrix. Furthermore,

$$\begin{aligned} DD' &= H_l Y (H_n \otimes H_m) Y' H_l \\ &= K_l K_l' Y (K_n K_n' \otimes K_m K_m') Y' K_l K_l', \end{aligned} \tag{3.43}$$

where K_l is the $l \times (l-1)$ semi-orthogonal matrix such that

$$K_l K_l' = H_l = I_l - \tfrac{1}{l} \underline{1}_l \underline{1}_l',$$

$$K_l' K_l = I_{l-1}, \text{ and } K_l' = K_l^{-},$$

where K_l^{-} is the Moore-Penrose inverse of K_l.

Thus it is easily seen (see Johnson and Graybill(1972)), that the nonzero eigenvalues of

DD' are the same as the nonzero eigenvalues of the full rank $(l-1)\times(l-1)$ matrix W where

$$W = K_l' DD' K_l$$

$$= K_l' Y (K_n K_n' \otimes K_m K_m') Y' K_l. \tag{3.44}$$

Notice that the role of K_l is to remove the singularity in DD' and make W a full rank matrix. The following theorem gives the distribution of the matrix W.

Theorem 3.4.3: The elements of the matrix W have a noncentral Wishart distribution, i.e. W is distributed as

$$W_{l-1}(\ (n-1)(m-1),\ \sigma^2 I_{l-1},\ K_l' \Gamma\Gamma' K_l\ ) \tag{3.45}$$

where $\Gamma = (\eta_{ijk})$.

Proof: The columns of $K_l' Y$, i.e., $K_l' \underline{y}_i$, are distributed independently as $(l-1)$ variate $N(K_l' M,\ \sigma^2 K_l' K_l)$, that is,

$$K_l' Y \overset{\text{indep.}}{\sim} N_{l-1}(K_l' M,\ \sigma^2 K_l' K_l)$$

where $M = (m_{ijk}) = E[Y],\ m_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \eta_{ijk}$

Since $(K_n K_n' \otimes K_m K_m')$ is a symmetric idempotent matrix with rank $(n-1)(m-1)$,

$$W = (K_l' Y)(K_n K_n' \otimes K_m K_m')(K_l' Y)'$$
$$\sim W_{l-1}(\ (n-1)(m-1),\ \sigma^2 I_{l-1},\ \Theta\ )$$

where the noncentrality parameter

$$\Theta = K_l' M (K_n K_n' \otimes K_m K_m') M' K_l = K_l' \Gamma\Gamma' K_l. \qquad \square$$

Hence the nonzero eigenvalues of the matrix DD' are distributed as the eigenvalues of a matrix W which has the Wishart distribution (3.45). Notice that, under the assumption $c^2 \simeq \lambda$, i.e., $\eta_{ijk} = c\, g_i h_j e_k \simeq \lambda\, g_i h_j e_k$, for all i, j, and k, then the non-centrality matrix $K_l' \Gamma \Gamma' K_l$ has rank one and hence there is only one nonzero eigenvalue $\hat{\lambda}_1$.

Distribution of L

Now the distribution of the statistic

$$L = \frac{\hat{\lambda}_1}{\sum_{p=1}^{l-1} \hat{\lambda}_p},$$

under the null hypothesis H_0: $\lambda_1 = 0$ is formulated. As stated above, the eigenvalues of W are $\hat{\lambda}_1 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_{l-1}$ and under the H_0: $\lambda_1 = 0$, W is distributed

$$W_{l-1}(\,(n-1)(m-1),\, \sigma^2 I_{l-1},\, 0).$$

The joint distribution of the eigenvalues is given by Seber (1984)

$$f(\,\hat{\lambda}_1, \cdots, \hat{\lambda}_{l-1}) = f(\,\underline{\hat{\lambda}}) = k\, \Big(\prod_{i=1}^{l-1} \hat{\lambda}_i\Big)^{\frac{(n-1)(m-1)-(l-1)-1}{2}} \prod_{i<j}^{l-1} (\hat{\lambda}_i - \hat{\lambda}_j)\; e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{l-1} \hat{\lambda}_i} \tag{3.46}$$

in the region for which $0 < \hat{\lambda}_{l-1} \leq \cdots \leq \hat{\lambda}_2 \leq \hat{\lambda}_1 < +\infty$, and k is the constant such that

$$k = \frac{\pi^{\frac{l-1}{2}}}{(2\sigma^2)^{\frac{(n-1)(m-1)-(l-1)-1}{2}} \prod_{i=1}^{l-1} \left[\Gamma\left(\frac{(l-1)-i+1}{2}\right) \Gamma\left(\frac{(n-1)\,(m-1)\,-i+1}{2}\right)\right]}.$$

To find the distribution of L, first make the following transformation. Define the transformation random variable T by

$$T_i = \frac{\lambda_i}{\sum_{p=1}^{l-1} \lambda_p}, \quad i = 1, \cdots, l-2 \qquad \text{and} \qquad T_{l-1} = \sum_{p=1}^{l-1} \lambda_p$$

and consider the corresponding transformations

$$t_i = \frac{\hat{\lambda}_i}{\sum_{p=1}^{l-1} \hat{\lambda}_p}, \quad i = 1, \cdots, l-2 \quad \text{and} \quad t_{l-1} = \sum_{p=1}^{l-1} \hat{\lambda}_p .$$

Let the $t_i = u_i(\hat{\lambda}_1, \cdots \hat{\lambda}_{l-1})$ and its inverse of t_i be denoted by $\hat{\lambda}_i = \omega_i(t_1, t_2, \cdots, t_{l-1})$, where, $i = 1, \cdots, l-1$. Thus the transformation variables and its respective inverses are

$u_i(\hat{\lambda}_1 + \cdots + \hat{\lambda}_{l-1})$	$\omega_i(t_1, t_2, \cdots, t_{l-1})$
$t_1 = \dfrac{\hat{\lambda}_1}{\hat{\lambda}_1 + \cdots + \hat{\lambda}_{l-1}}$	$\hat{\lambda}_1 = t_1 t_{l-1}$
$t_2 = \dfrac{\hat{\lambda}_2}{\hat{\lambda}_1 + \cdots + \hat{\lambda}_{l-1}}$	$\hat{\lambda}_1 = t_2 t_{l-1}$
$\vdots$	$\vdots$
$t_{l-2} = \dfrac{\hat{\lambda}_{l-2}}{\hat{\lambda}_1 + \cdots + \hat{\lambda}_{l-1}}$	$\hat{\lambda}_{l-2} = t_{l-2} t_{l-1}$
$t_{l-1} = \hat{\lambda}_1 + \cdots + \hat{\lambda}_{l-1}$	$\hat{\lambda}_{l-1} = (1 - t_1 - \cdots - t_{l-2})\, t_{l-1}$.

The space $\wp$ is the $l-1$ dimensional region of the ranges of the λ's $0 < \hat{\lambda}_{l-1} \leq \cdots \leq \hat{\lambda}_2 \leq \hat{\lambda}_1 \leq +\infty$. The one-to-one transformation $t_i = u_i(\hat{\lambda}_1, \cdots \hat{\lambda}_{l-1})$ maps a set $\wp$ in the $l-1$ dimensional space onto a set of the ß space, which consists of the following ranges of the t_i's:

$$x_1 = \frac{1}{l-1} < t_1 < 1 = y_1$$

$$x_2 = \frac{1-t_1}{l-2} < t_2 < \min(t_1, 1-t_1) = y_2$$

$$x_3 = \frac{1-t_1-t_2}{l-3} < t_3 < \min(t_2, 1-t_1-t_2) = y_3$$

$$x_4 = \frac{1-t_1-t_2-t_3}{l-4} < t_4 < \min(t_3, 1-t_1-t_2-t_3) = y_4$$

$$\vdots$$

$$x_{l-2} = \frac{1-t_1-t_2-\cdots-t_{l-3}}{2} < t_{l-2} < \min(t_{l-3},\ 1-t_1-t_2-\cdots-t_{l-3}) = y_{l-2}, \text{ and}$$

$$x_{l-1} = 0 < t_{l-1} < \infty = y_{l-1}. \tag{3.47}$$

It will be assumed that the the first-order partial derivatives are continuous and that the Jacobian J is not identically zero in ß. The Jacobian of the transformation is $(t_{l-1})^{l-2}$ as shown below:

$$J_{(l-1)\times(l-1)} = \begin{vmatrix} \frac{\partial \hat{\lambda}_1}{\partial t_1} & \frac{\partial \hat{\lambda}_1}{\partial t_2} & \cdots & \frac{\partial \hat{\lambda}_1}{\partial t_{l-1}} \\ \frac{\partial \hat{\lambda}_2}{\partial t_1} & \frac{\partial \hat{\lambda}_2}{\partial t_2} & \cdots & \frac{\partial \hat{\lambda}_2}{\partial t_{l-1}} \\ & \vdots & & \\ \frac{\partial \hat{\lambda}_{l-1}}{\partial t_1} & \frac{\partial \hat{\lambda}_{l-1}}{\partial t_2} & \cdots & \frac{\partial \hat{\lambda}_{l-1}}{\partial t_{l-1}} \end{vmatrix}$$

$$= \det \begin{bmatrix} t_{l-1} & 0 & 0 & 0 & t_1 \\ 0 & t_{l-1} & 0 & 0 & t_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & t_{l-1} & t_{l-2} \\ -t_{l-1} & -t_{l-1} & \cdots & -t_{l-1} & (1-t_1-\cdots-t_{l-2}) \end{bmatrix},$$

now by adding all the rows to the $(l-1)$th row, we obtain

$$= \det \begin{bmatrix} t_{l-1} & 0 & 0 & 0 & 0 \\ 0 & t_{l-1} & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & t_{l-1} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= (t_{l-1})^{l-2}.$$

Before the joint density of $(t_1, \cdots, t_{l-1})$ is given, rewrite the density $f(\hat{\lambda}_1, \cdots, \hat{\lambda}_{l-1})$ in (3.46) as

$$f(\underline{\hat{\lambda}}) = k \left(\prod_{i=1}^{l-2} \hat{\lambda}_i \right)^{\frac{(n-1)(m-1)-l}{2}} (\hat{\lambda}_{l-1})^{(n-1)(m-1)-l} \prod_{i<j}^{l-2} (\hat{\lambda}_i - \hat{\lambda}_j) \prod_{i=1}^{l-2} (\hat{\lambda}_i - \hat{\lambda}_{l-1}) \; e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{l-1} \hat{\lambda}_i}$$

for the range given by (3.47). The joint density of $T_1, \cdots, T_{l-1}$ is then

$g(t_1, \cdots, t_{l-1})$

$= |J| \times f(\hat{\lambda}_1, \cdots, \hat{\lambda}_{l-1})$

$$= J \times f(\, t_1 t_{l-1},\, t_2 t_{l-1},\, \cdots,\, t_{l-2} t_{l-1},\, (1-t_1- \cdots -t_{l-2}) t_{l-1}\,)$$

$$= k\, (t_{l-1}^{\frac{(n-1)(m-1)-l}{2}})^{l-2}\, (\prod_{i=1}^{l-2} t_i)^{\frac{(n-1)(m-1)-l}{2}}\, (t_{l-1})^{\frac{(n-1)(m-1)-l}{2}}$$

$$\times\, (1-t_1- \cdots -t_{l-2})^{\frac{(n-1)(m-1)-l}{2}} \quad \prod_{i<j}^{l-2} \{(t_i-t_j)\, t_{l-1}\} \quad \prod_{i=1}^{l-2} \{[t_i- (1-t_1- \cdots -t_{l-2})]\, t_{l-1}\}$$

$$\times\, e^{-\frac{1}{2\sigma^2}\, [\sum_{i=1}^{l-2} t_i t_{l-1} \,+(1-t_1 \cdots t_{l-2})\, t_{l-1}]} \,.$$

Since the summation part of the exponent term becomes t_{l-1}, and we can rewrite the following expression on the second line

$$\prod_{i<j}^{l-2} \{(t_i-t_j)\, t_{l-1}\} \quad \prod_{i=1}^{l-2} \{[t_i- (1-t_1- \cdots -t_{l-2})]\, t_{l-1}\}$$

as

$$(t_{l-1})^{\frac{(l-1)(l-2)}{2}+(l-2)} \quad \prod_{i<j}^{l-2} (t_i-t_j) \quad \prod_{i=1}^{l-2} (t_i-1+t_1+ \cdots +t_{l-2}),$$

and we have the joint probability density function

$g(t_1, \cdots, t_{l-1})$

$$= \frac{\pi^{\frac{l-1}{2}}\, (\, t_{l-1}^{\frac{(n-1)(m-1)(l-1)}{2}-1}\,)\, e^{-\frac{t_{l-1}}{2\sigma^2}}}{(2\sigma^2)^{\frac{(n-1)(m-1)(l-1)-1}{2}} \prod_{i=1}^{l-1} [\Gamma(\frac{(l-1)-i+1}{2})\, \Gamma(\frac{(n-1)\,(m-1)\,-i+1}{2})]}$$

$$[\,(1-t_1-\cdots-t_{l-2})\prod_{i=1}^{l-2}t_i\,]^{\frac{(n-1)(m-1)-l}{2}}\prod_{i<j}^{l-2}(t_i-t_j)\prod_{i=1}^{l-2}(t_i-1+t_1+\cdots+t_{l-2})$$

(3.48)

for the region ß in (3.47).

From (3.48), we see that T_{l-1} is independent of $(T_1, \cdots, T_{l-2})$; in particular, T_{l-1} is distributed as the gamma distribution

$$\Gamma\left(\frac{(n-1)(m-1)(l-1)}{2},\ 2\sigma^2\right)$$

for $0 < t_{l-1} < \infty$. In other words, apart from σ^2, the random variable T_{l-1} has a chi-square distribution with $(n-1)(m-1)(l-1)$ degrees of freedom, i.e.,

$$T_{l-1} \sim \sigma^2\chi^2_{(n-1)(m-1)(l-1)}.$$

Subsequently, the joint density of $(T_1, \cdots, T_{l-2})$ is

$$g^*(t_1, \cdots, t_{l-2})$$

$$= k^*\,[\,(1-t_1-\cdots-t_{l-2})\prod_{i=1}^{l-2}t_i\,]^{\frac{(n-1)(m-1)-l}{2}}\prod_{i<j}^{l-2}(t_i-t_j)\prod_{i=1}^{l-2}(t_i-1+t_1+\cdots+t_{l-2})$$

(3.49)

where,

$$k^* = \frac{\pi^{\frac{l-1}{2}}\,\Gamma\left(\frac{(n-1)(m-1)(l-1)}{2}\right)}{\prod_{i=1}^{l-1}\left[\Gamma\left(\frac{(l-1)-i+1}{2}\right)\Gamma\left(\frac{(n-1)\,(m-1)\,-i+1}{2}\right)\right]},$$

for the range given for $(t_1, \cdots, t_{l-2})$ in (3.47).

Finally, we need the marginal density of T_1 since we want to find the distribution of

$$T_1 = \frac{\lambda_1}{\sum_{i=1}^{l-1} \lambda_i}.$$

From (3.49), we see that the marginal density is not depend on parameters. So the marginal density is T_1 computed as

$$g_1(t_1) = \int_{x_2}^{y_2} \int_{x_3}^{y_3} \cdots \int_{x_{l-1}}^{y_{l-2}} g^*(t_1, \cdots, t_{l-2}) \quad dt_{l-2} \cdots dt_3 \, dt_2 \tag{3.50}$$

for $\frac{1}{l-1} < t_1 < 1$.

Notice that, apart from some parameters, the distributions of M(3.28) and N(3.29) are similar to the distribution of T_1 and therefore to the distribution of L(3.27) (since T_1 and L are the same).

Remarks

Now what we need is the table of critical points of T_1. It may be computed possibly by

(i) numerically computing the density in (3.50). For this approach, the question arises as to its efficiency compared with other methods, such as, randomization tests in section 3.3

(ii) approximating the distribution using, for example, Dirichlet or Beta. Notice that for the special case of one random variable, the Dirichlet p.d.f. becomes a beta p.d.f.

(iii) applying large sample asymptotic distribution theory (a χ^2 distribution). Note that for this application, the basic asymptotic regularity conditions (Rao, section 6a) needs to be checked.

(iii) using Monte Carlo methods similar to the studies in Mandel (1970).

So far we studied the test of a single multiplicative term to represent the three-factor interaction. Although the key statistics we consider are $\hat{c}^2 = \hat{c}^2_{(1)}$ (3.15) or $\hat{c}^2_{111}$, there is no reason why more than one should not be used at the same time. When more than one term is to be considered to separate interaction from random error, a criterion is needed so that an appropriate number of terms can be selected to describe interaction. One way to deal with this issue is to perform stepwise procedures based on testing procedures discussed in this chapter and in Appendix C. The approach is similar to the method presented by Yochmowitz and Cornell (1978) for two-way tables. An alternative approach based on Monte Carlo methods is discussed in the following section.

3.5 MONTE CARLO METHODS

Monte Carlo methods are often used by statisticians to investigate distributional problems that are mathematically intractable, such as evaluation of distribution functions or moments of a distribution (Hartley, 1977). As stated earlier, the distribution of the estimated squared core elements c^2_{pqr} is unknown. To overcome this, in Section 3.3, the distribution of these variables are studied by applying randomization tests which can be thought of as special cases within a broader category of Monte Carlo tests.

Consider, a three-way fixed effects ANOVA model (3.5) with one observation per cell with more than a single multiplicative term:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \sum_{p=1}^{s}\sum_{q=1}^{t}\sum_{r=1}^{u} c_{pqr}\, g_{ip}\, h_{jq}\, e_{kr} + \epsilon^*_{ijk}.$$

When more than one term is to be used to differentiate interaction from random error, one of the main problems is to derive a criterion by which a decision can be made for selecting an appropriate number of terms to describe interaction. In this section, Monte Carlo methods will be applied to determine the appropriate number of multiplicative terms (s, t, u) that would summarize the interaction structure. This application is an extension of Mandel's (1971) approach for an unreplicated two-way model.

3.5.1 ANOVA table with multiplicative terms

First, an ANOVA table based on the model (3.5) is constructed as shown in Table 9 where the three-factor interaction term is partitioned into multiplicative terms using three-mode principal component analysis. In addition to the usual rows, for an ANOVA table for the model with no three-factor interaction term (3.2), it displays several rows with terms $\hat{c}_{pqr}\hat{g}_{ip}\hat{h}_{jq}\hat{e}_{kr}$, $1 \le p \le l$, $1 \le q \le m$, $1 \le r \le n$. These additional terms represent the corresponding three-factor interaction terms. To follow the usual representation of ANOVA, we need to find a number of degrees of freedom, a sum of squares, and a mean square for the corresponding multiplicative term.

The sums of squares corresponding to the multiplicative terms are, simply the quantities $\hat{c}^2_{pqr}$, $1 \le p \le l$, $1 \le q \le m$, and $1 \le r \le n$. The reason for this is that the $\hat{c}^2_{pqr}$ are additive components in the decomposition of the residual sum of squares

Table 9. Analysis of variance for the multiplicative model (3.5)

Source	df	SS	MS
		usual	usual
α_i	$(l-1)$	SS(A)	MS(A)
β_j	$(m-1)$	SS(B)	MS(B)
γ_k	$(n-1)$	SS(C)	MS(C)
$(\alpha\beta)_{ij}$	$(l-1)(m-1)$	SS(AB)	MS(AB)
$(\alpha\gamma)_{ik}$	$(l-1)(n-1)$	SS(AC)	MS(AC)
$(\beta\gamma)_{jk}$	$(m-1)(n-1)$	SS(BC)	MS(BC)
η_{ijk}	$(l-1)(m-1)(n-1)$	SS(ABC)	MS(ABC)
$c_{111}g_{i1}h_{j1}e_{k1}$	δ_{111}	$\hat{c}^2_{111}$	$\hat{c}^2_{111}/\delta_{111}$
$c_{112}g_{i1}h_{j1}e_{k2}$	δ_{112}	$\hat{c}^2_{112}$	$\hat{c}^2_{112}/\delta_{112}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$c_{pqr}g_{ip}h_{jq}e_{kr}$	δ_{pqr}	$\hat{c}^2_{pqr}$	$\hat{c}^2_{pqr}/\delta_{pqr}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$c_{stu}g_{is}h_{jt}e_{ku}$	δ_{stu}	$\hat{c}^2_{stu}$	$\hat{c}^2_{stu}/\delta_{stu}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$

Notations: (a) $\hat{c}^2_{pqr}$ is the sum of squares for (p,q,r) th entry and

(b) δ_{pqr} is the corresponding number of degrees of freedom.

$$SS(ABC) = \sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}^2 = \sum_{p=1}^{s}\sum_{q=1}^{t}\sum_{r=1}^{u} \hat{c}_{pqr}^2 \qquad (3.51)$$

in the case of a complete decomposition, i.e., $s = l$, $t = m$, and $u = n$. The second equality in (3.51) follows from (3.7) and the orthogonality property of eigenvectors $\underline{g}$, $\underline{h}$, and $\underline{e}$, by which all cross products are zero. The detailed proof for this is given in Appendix A. From the relation in (3.51), one can see that the squared generalized singular value, c_{pqr}^2, indicates how the sum of squares for three-factor interaction, SS(ABC), or the total variation of the three-factor interaction, is accounted for by this particular combination of components. Furthermore the core matrix partitions the fitted variation into parts which can be attributed to each component. In particular, the equation suggests that the interaction sum of squares is partitioned into $s \times t \times u$ terms $\hat{c}_{111}^2, \cdots, \hat{c}_{stu}^2$. In practice, only the first few largest terms are retained for the interaction sum of squares and the other terms are pooled as an error term.

Degrees of freedom are denoted as δ_{pqr} as in Table 9. As stated earlier, $\hat{c}_{pqr}^2$ are neither distributed independently nor as chi-square variables. Therefore, the degrees of freedom, δ_{pqr}, cannot be obtained analytically. These will be obtained using Monte Carlo methods in this section. Before we proceed, the definition of the degrees of freedom is given first. Consider the last column, the "MS" column, in Table 9. The number of degrees of freedom, δ_{pqr}, is the divisor corresponding to $\hat{c}_{pqr}^2$. In the absence of real systematic interaction effects, the ratio $\hat{c}_{pqr}^2/\delta_{pqr}$ is merely an estimate of random experimental error. Thus if $\sigma^2 = \mathrm{var}(\epsilon_{ijk})$, and if none of the terms $c_{pqr}g_{ip}h_{jq}e_{kr}$ occur in the real model, then we must have

$$E\left[\frac{\hat{c}_{pqr}^2}{\delta_{pqr}}\right] = \sigma^2 \quad \text{for each} \quad pqr,$$

where $\hat{c}_{pqr}^2$ is the squared core element. Or equivalently, the number of degrees of freedom can be defined as

$$\delta_{pqr} = \frac{E[\hat{c}_{pqr}^2]}{\sigma^2}. \qquad (3.52)$$

That is, apart from σ^2, these degrees of freedom are based on the expected values of the squared core elements $\hat{c}^2_{pqr}$.

As stated, the degrees of freedom, δ_{pqr}, will be obtained using simulation methods. In particular, Monte Carlo methods will be applied to compute the expectation in (3.52) to find the degrees of freedom. Using these values of δ_{pqr}, one can then examine the magnitude of $\hat{c}^2_{pqr}/(\hat{\sigma}^2\delta_{pqr})$ to determine which of the $\hat{c}^2_{pqr}$'s are significant. Here $\hat{\sigma}^2$ is the proportion of the sum of all the remaining $\hat{c}^2_{pqr}$, i.e., those with insignificant magnitude, with respect to the corresponding degrees of freedom. So, for example, if $\hat{c}^2_{111}$ and $\hat{c}^2_{112}$ are considerably larger than other elements then the estimated variance is

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d^2_{ijk} - \hat{c}^2_{111} - \hat{c}^2_{112}}{\sum_{p=1}^{l}\sum_{q=1}^{m}\sum_{r=1}^{n}\delta_{pqr} - \delta_{111} - \delta_{112}}$$

$$= \frac{SS(ABC) - \hat{c}^2_{111} - \hat{c}^2_{112}}{(l-1)(m-1)(n-1) - \delta_{111} - \delta_{112}}.$$

3.5.2 Application of Monte Carlo Methods

In this section, a procedure of Monte Carlo application is discussed through an example to illustrate how to select a reasonable number of terms to represent the interaction.

To perform the analysis, one can start with an $l \times m \times n$ array of independent, normally distributed deviates, d_{ijk}, with common variance σ^2 and having no three-factor interaction. It is necessary to repeat the analysis a large number of times, taking a different set of random normal deviates each time and averaging, for each value of pqr

separately, the $\hat{c}^2_{pqr}$ values obtained in all sets.

The procedure for Monte Carlo estimation of degrees of freedom is summarized as follows:

1. Generate a different set of random normal deviates, d_{ijk}, with $\sigma^2 = 1$, without loss of generality, i.e., d_{ijk} are from $N(0,1)$

2. For each generated data in step 1, the three-mode principal component analysis is applied to partition the deviates, and obtain the squared core elements as the sum of squares for each interaction term.

3. Repeat steps 1 and 2 a large number of times, taking different sets of residuals.

4. Compute the mean of $\hat{c}^2_{pqr}$, which is the degrees of freedom δ_{pqr} for the (p,q,r)th term, and the standard deviation of the corresponding term.

5. Construct an ANOVA table similar to Table 9 with $s \times t \times u$ multiplicative terms. These multiplicative terms are entered in a decreasing order with respect to the values of the sum of squares or the squared core elements.

6. To select a number of multiplicative term(s), first compute the following F-like statistic:

$$\frac{\text{MS}(c_{pqr}\text{ term})}{\text{MS(remainder term)}} = \frac{\hat{c}^2_{pqr}}{\text{MS(remainder term)}\ \delta_{pqr}}, \tag{3.53}$$

where,

$$\text{MS(remainder term)} = \frac{\text{SS(ABC)} - \sum_{p=1}^{s}\sum_{q=1}^{t}\sum_{r=1}^{u}\hat{c}^2_{pqr}}{\text{d.f.(ABC)} - \sum_{p=1}^{s}\sum_{q=1}^{t}\sum_{r=1}^{u}\delta_{pqr}}. \tag{3.54}$$

If the ratio is considerably larger than 1, the particular term should be included in the model to represent the interaction. On the other hand, if the ratio is quite small, i.e., near or less than 1, then the term is considered to be insignificant. These terms with insignificant values of the ratio are pooled together and used as experimental error.

7. Finally, write a suitable model based on the results from the previous steps and reconstruct an ANOVA table with an appropriate number of multiplicative terms.

Example

The principle behind a Monte Carlo application of the type just described is now demonstrated for Example 1 in Section 3.3.4. The dataset for the example is displayed in Table 1 and the model for the data is given in the section. The aim of this illustration is to show how to select an appropriate number of multiplicative terms to describe three-factor interaction using a Monte Carlo method.

Following the steps above, a Monte Carlo experiment is carried out using a SAS/IML program. The procedure generates 1200 independent sets of normal errors, d_{ijk} ($\mathrm{d}_{ijk} \sim N(0,1)$), for the combination of $l = 6$, $m = 5$, and $n = 4$. These are then viewed as residuals from a three-way factorial experiment with no three-factor interaction. Of course, one can construct tables of many different combination of l, m, and n in a similar manner.

First, for each independent set of deviates, the three-mode principal component analysis was fitted. From the analysis, core elements are obtained based on the $s = t = u = 2$ reduction. Then, for each combination of (p,q,r), the degrees of freedom which is the average of the statistics $\hat{\mathrm{c}}^2_{pqr}$ and their standard deviations are calculated.

The first row of Table 10 gives the observed sums of squares (or squared core elements) as explained in Section 3.5.1, and the second row shows its proportion relative to SSE,

or SS(ABC) in this case, which is obtained by fitting the model (3.2). The third row gives the numbers of degrees of freedom, δ_{pqr}, which are the averages of the sum of squares. Their standard deviation is displayed in the last row of the table. In addition to the $s \times t \times u = 8$ interaction sum of squares terms, the table displays the results for the statistic $c^2_{(1)}$, the maximum core elements (3.15).

Table 10. Values of $\hat{c}^2_{pqr}$, δ_{pqr}, and their standard deviations for, $l = 6$, $m = 5$, and $n = 4$.

	$c^2_{(1)}$	c^2_{111}	c^2_{222}	c^2_{212}	c^2_{221}	c^2_{122}	c^2_{121}	c^2_{112}	c^2_{211}
Observed	2893.36	2893.36	13.99	5.29	3.96	0.18	0.03	0.02	0.002
$\frac{c^2_{pqr}}{SS(ABC)}$	0.978	0.978	0.005	0.002	0.001	< .001	< .001	< .001	< .001
Mean(δ_{pqr})	19.20	15.31	2.30	3.79	4.52	3.51	5.97	5.79	6.28
SD	6.57	10.11	3.58	4.08	4.53	3.66	5.58	5.52	6.02

As the first step in our solution, a complete analysis of variance table, which is similar to Table 9, is constructed for Example 1. Table 11 displays, in addition to the usual seven terms, several more items corresponding to the partitioning of the interaction term. Following the usual pattern of the analysis of variance, each of these additional terms contain the degrees of freedom, a sum of squares, and a mean square. The corresponding degrees of freedom are from the third row of Table 10. The sum of squares are the squared core elements. Mean squares for these multiplicative terms are obtained by dividing the sum of squares by the corresponding degrees of freedom, δ_{pqr}, as usual. The remainder term is obtained by taking difference.

Table 11. Analysis of variance with multiplicative terms for Example 1

Source	Terms in model	df	SS	MS	F	P-value
A	α_i	5	1482.2198	296.4440	6.01	0.0001
B	β_j	4	793.5262	198.3816	4.02	0.0059
C	γ_k	3	398.4869	132.8290	2.69	0.0540
A*B	$(\alpha\beta)_{ij}$	20	9720.3917	486.0196	9.86	0.0001
A*C	$(\alpha\gamma)_{ik}$	15	20.7295	1.3820	0.03	1.0000
B*C	$(\beta\gamma)_{jk}$	12	11.1497	0.9291	0.02	1.0000
A*B*C	η_{ijk}	60	2958.284	49.305		
	$c_{111}g_{i1}h_{j1}e_{k1}$	15.3	2893.36	189.11		
	$c_{222}g_{i2}h_{j2}e_{k2}$	2.3	13.99	6.08		
	$c_{212}g_{i2}h_{j1}e_{k2}$	3.8	5.29	1.39		
	$c_{221}g_{i2}h_{j2}e_{k1}$	4.5	3.96	0.88		
	$c_{122}g_{i1}h_{j2}e_{k2}$	3.5	0.18	0.06		
	$c_{121}g_{i1}h_{j2}e_{k1}$	5.9	0.03	0.005		
	$c_{112}g_{i1}h_{j1}e_{k2}$	5.8	0.02	0.0034		
	$c_{211}g_{i2}h_{j2}e_{k1}$	6.3	0.002	0.0003		
Remainder	ϵ^*_{ijk}	60-47.4 = 12.6	2958.28-2916.83 = 41.45	3.30		
Total		119	15384.78			

From Table 11, it is noted that the interaction can be summarized by the first multiplicative term. The mean squares for the first term is significantly larger than other terms. This term alone accounts for 97.8% of the sum of squares of the three-factor interaction, while the other elements contributes less than 3% of its variation. In fact, the F-like ratio (3.53)

$$\frac{\text{MS}(c_{111}\text{ term})}{\text{MS(remainder term)}} = \frac{189.11}{3.30} = 57.31$$

for the first term while the ratios for other terms are less than 1.9. Therefore all the terms, except the first term, are pooled together and considered as the residual term in a reconstructed ANOVA table (Table 12).

Table 12 shows the revised version of Table 11. It partitions the interaction term into a single multiplicative term and a residual. Notice that the mean squares for the first term (= 189.11) is considerably larger than the pooled error variance (= 1.45). Hence, one may conclude that the model for the data need a three three-factor interaction term and it is summarized well by single multiplicative term.

Thus combining these results, only the first multiplicative term seems to be necessary in order to explain the variability of three-factor interaction in this data. So one may conclude the fitted model for the data is

$$\hat{y}_{ijk} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + (\widehat{\alpha\beta})_{ij} + (\widehat{\alpha\gamma})_{ik} + (\widehat{\beta\gamma})_{jk} + \hat{c}_{111}\,\hat{g}_{i1}\,\hat{h}_{j1}\,\hat{e}_{k1},$$

where $\hat{c}_{111} = -54.30$ and $\underline{\hat{g}}'_{(1)} = (\ -.47\ -.36\ -.24\ -.00\ .47\ .61)$, $\underline{\hat{h}}'_{(1)} = (\ -.51\ -.35\ -.18\ .36\ .68)$, and $\underline{\hat{e}}'_{(1)} = (\ -.43\ -.21\ .62\ .62)$ are the first column vectors of the orthonormal component matrices G, H, and E respectively.

Table 12. Analysis of variance with single multiplicative term for Example 1

Source	Terms in model	df	SS	MS	F	P-value
A	α_i	5	1482.2198	296.4440	204.44	<0.0001
B	β_j	4	793.5262	198.3816	136.81	<0.0001
C	γ_k	3	398.4869	132.8290	91.61	<0.0001
A*B	$(\alpha\beta)_{ij}$	20	9720.3917	486.0196	335.19	<0.0001
A*C	$(\alpha\gamma)_{ik}$	15	20.7295	1.3820	0.95	0.5170
B*C	$(\beta\gamma)_{jk}$	12	11.1497	0.9291	0.64	0.7956
A*B*C	η_{ijk}	60	2958.284	49.305		
	$c_{111}g_{i1}h_{j1}e_{k1}$	15.3	2893.36	189.11	130.42	<0.0001
Residual	ϵ^*_{ijk}	44.7	64.92	1.45		
Total		119	15384.78			

The following two remarks are worth to mention:

1. Before the analysis is concluded, as always, a residual analysis is suggested to make sure that the residuals from a new model with multiplicative term(s) show no systematic pattern.

2. In order to learn about the interaction structure, one needs to study the elements of the generalized singular values, i.e., core elements, and their corresponding the eigenvectors.

Remarks

This chapter is concluded with Table 13, which presents some feasible models with at least one multiplicative interaction term using the three-mode PCA model. For this presentation, write the model (3.3) as

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \xi_{ijk} + \epsilon_{ijk} \tag{3.55}$$

where ξ_{ijk} represents the interactions.

Model 1 in Table 13 is the three-factor additive model. Model 2 decomposes the two-plus-three way interaction terms. Model 3 partitions only the three-factor interaction term, whereas, for the same model, model 4 decomposes both the two-factor (using SVD) and the three-factor interaction terms. These models are more appropriate for the two-factor ANOVA model with blocking i.e., the three-factor randomized block design. Lastly, model 5 denotes the unreplicated three-factor ANOVA model with three-way interaction term decomposed as in model (3.5).

For multi-level analysis of variance, the nature and the structure of interactions are

often difficult to interpret. More insight can be obtained by structuring data with a multiplicative model. Thus, the method presented in this section together with the method in Section 3.3 will be useful not only for the nonreplicated experiments but also for the usual replicated factorial experiments.

Table 13. Some multiplicative models for three-way tables

Model	Interaction form
Model 1	$\xi_{ijk} = 0$
Model 2	$\xi_{ijk} = \sum_{p=1}^{s}\sum_{q=1}^{t}\sum_{r=1}^{u} c_{pqr}\, g_{ip}\, h_{jq}\, e_{kr}$
Model 3	$\xi_{ijk} = (\alpha\beta)_{ij} + \sum_{p=1}^{s}\sum_{q=1}^{t}\sum_{r=1}^{u} c_{pqr}\, g_{ip}\, h_{jq}\, e_{kr}$
Model 4	$\xi_{ijk} = \sum_{p=1}^{s} \lambda_p u_{pi} v_{pj} + \sum_{p=1}^{s}\sum_{q=1}^{t}\sum_{r=1}^{u} c_{pqr}\, g_{ip}\, h_{jq}\, e_{kr}$
Model 5	$\xi_{ijk} = (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \sum_{p=1}^{s}\sum_{q=1}^{t}\sum_{r=1}^{u} c_{pqr}\, g_{ip}\, h_{jq}\, e_{kr}$

CHAPTER 4. GRAPHICAL PROCEDURES IN MODEL BUILDING

4.1 INTRODUCTION

In this chapter, a graphical procedure is discussed for diagnosing the type of model to fit to three-way tables. This is done by studying the relationship between various models and the geometrical configurations using joint plots from three-mode principal component analysis.

Several methods for diagnosing a model of a two-factor data matrix by means of graphical tools have been introduced by many authors, for instance, Snee (1982) used interaction plots (see also Milliken and Johnson (1989)), Bradu and Gabriel (1978) used the biplot and Kester (1979) studied possible application of Bradu and Gabriel's method to the multifactor concurrent model.

Gabriel (1971) introduced the two-dimensional biplot for a data matrix as a graphical method for displaying multiplicative effects in a single diagram. Bradu and Gabriel (1978) showed that the biplot of a two-way table can be used to diagnose the suitability of various special forms of the multiplicative model $y_{ij} = \mu + \alpha_i + \beta_j + \alpha_i^* \beta_j^*$. Gower (1990) generalized the method to a rank three data-matrix. He discussed the geometric configurations of three-dimensional biplots for data which fit some special forms of the

model $y_{ij} = \mu + \alpha_i + \beta_j + \alpha_i^* \beta_j^* + \alpha_i^{**} \beta_j^{**}$, including the full biplot of the previous model. In a different context, Gower and Digby (1981) and Gower and Harding (1988) discussed the relationship between the biplot representation and principal component analysis.

The joint plots of a three-way table may be viewed as a multi-way analogue of the biplot, providing a graphical representation of the data. In this chapter it is proposed that joint plots can be used as a data analytic tool for diagnosing the type of model to fit to three-way data. The joint plot has been introduced by Kroonenberg as an analogue of the biplot (which is associated only with two-factor layouts), but it has not been used as a tool for diagnosing the models for three-factor experiments in the literature. (It is not known to us at least).

Some models in the three-factor case may be diagnosed by checking whether each of the interactions is of a particular form. This type of diagnosis may be done by using interaction plots and biplots after reducing the three-factor array to a suitable matrix form called the bistructure as discussed in sections 4.2 and 4.3. Sometimes, however entire models of certain special forms can be diagnosed as a whole by using the three-mode PCA method and the corresponding joint plots. This allows a simpler procedure than that entailed by separate diagnosis of each interaction. In section 4.4, it will be discussed in detail.

This chapter is organized as follows: Sections 4.2 and 4.3 presents background materials on the interaction plots and the biplot method. The joint plots and our main results will be discussed in section 4.4 which consists of three subsections. First, the definition of joint plots is given in section 4.4.1 and a list of special models, each of whose joint plot will correspond to a special geometric configuration, is proposed in 4.4.2. Next, in section 4.4.3, a conversion table is given to see the clear relation between the parameters and bistructure representations coming from the graphical procedure. The main results are discussed in 4.4.4. An example is given in 4.5, to illustrate how to diagnose a model using the graphical technique and to show the usefulness of the method.

4.2 INTERACTION PLOTS

As stated in section 2.1, for an unreplicated two-way experiments, the following special cases of the general multiplicative model in Table 14, were studied, for instance, by Tukey(1949), Gollob(1968), and Mandel (1969).

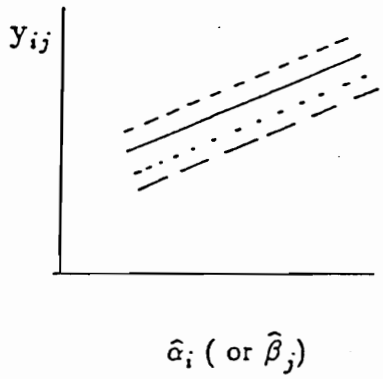
Table 14. Some multiplicative models for two-way tables

(a)	Additive model	$y_{ij} = \mu + \alpha_i + \beta_j$
(b)	Tukey's concurrent model	$y_{ij} = \mu + \alpha_i + \beta_j + \lambda\alpha_i\beta_j$
(c)	Mandel's rows-linear model	$y_{ij} = \mu + \alpha_i + \beta_j + \delta\alpha_i^*\beta_j$
(d)	Mandel's columns-linear model	$y_{ij} = \mu + \alpha_i + \beta_j + \xi\alpha_i\beta_j^*$

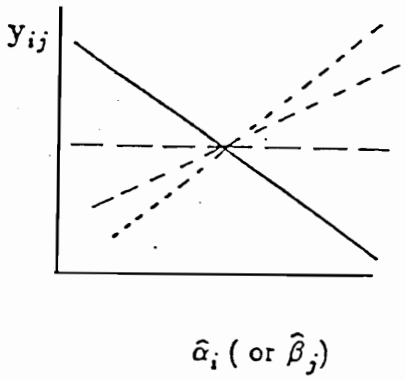
Note that $\alpha_i^* \neq \alpha_i$, $\beta_j^* \neq \beta_j$. In (b), interaction is modeled as a product of row and column effects while in (c) (and (d)), interaction is a product of the column (row) effect and a different row(column) effect.

Interaction plots for two-way tables are usually obtained by graphing $\hat{\mu}_{ij}$ versus a treatment level. Milliken and Johnson (1989) discuss on alternative type of interaction plots called type II interaction plots. The type II interaction plots for a two-way model are the plot of the points $(\hat{\alpha}_i, y_{ij})$ for $i = 1,\cdots,l$ for every $j = 1,\cdots,m$ or the plot of points $(\hat{\beta}_j, y_{ij})$ for every $i = 1,\cdots,l$. The plots are useful when we choose an appropriate model, especially if the type of model is going to be close to one of the above special forms. The

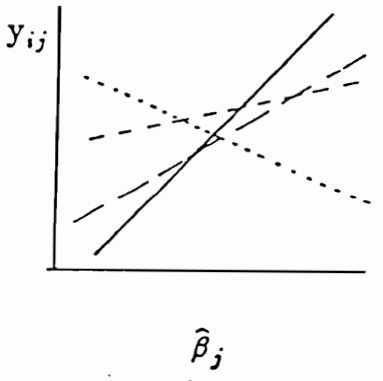
a. Additive model



b. Tukey's concurrent model



c. Mandel's rows-linear model



d. Mandel's columns-linear model

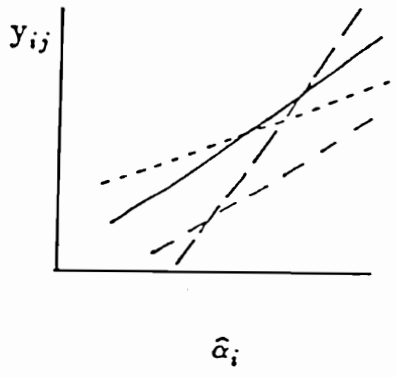


Figure 7. Interaction plots

sketch of interaction plots for the above special forms are displayed in Figure 7. These type of plots can also be used for the assessment of an certain models for three-way data with interaction. First, if the array fits the three-way additive model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k$$

then the interaction plot y_{ijk} *vs* $\hat{\alpha}_i$ is the collection of mn straight lines which are parallel. The same graphical display will be shown for data plots y_{ijk} *vs* $\hat{\beta}_j$ or y_{ijk} *vs* $\hat{\gamma}_k$, in this case.

For the concurrent model (Kester, 1979)

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \lambda(\alpha_i\beta_j + \beta_j\gamma_k + \alpha_i\gamma_k) + \lambda^2\alpha_i\beta_j\gamma_k, \tag{4.1}$$

again the plot y_{ijk} *vs* $\hat{\alpha}_i$ will consist of mn straight lines which meet at a single point. The reason is that the model can be also expressed as

$$y_{ijk} = \mu - \tfrac{1}{\lambda} + \lambda^2 \left(\alpha_i + \tfrac{1}{\lambda}\right)\left(\beta_j + \tfrac{1}{\lambda}\right)\left(\gamma_k + \tfrac{1}{\lambda}\right). \tag{4.1a}$$

Thus for fixed j and k, the relation between α_i and y_{ijk} will be a linear relation so that it appears to be a straight line in the coordinate plane with the slope $\lambda^2 \left(\beta_j + \frac{1}{\lambda}\right)\left(\gamma_k + \frac{1}{\lambda}\right)$ and vertical intercept $\mu - \frac{1}{\lambda} + \lambda \left(\beta_j + \frac{1}{\lambda}\right)\left(\gamma_k + \frac{1}{\lambda}\right)$. Also the line passes through the point $\left(-\frac{1}{\lambda}, \mu - \frac{1}{\lambda}\right)$ for all pairs j and k.

In general, for n-way data in an $l_1 \times l_2 \times \cdots \times l_n$ array, the concurrent model would be expressed as

$$y_{i_1 i_2 \cdots i_n} = \mu + \sum_{t=1}^{n} \alpha_{i_t}^{(t)} + \lambda \sum_{\{s,t\} \in \wp_2} \alpha_{i_s}^{(s)} \alpha_{i_t}^{(t)} + \lambda^2 \sum_{\{s,t,u\} \in \wp_3} \alpha_{i_s}^{(s)} \alpha_{i_t}^{(t)} \alpha_{i_u}^{(u)}$$

$$+ \cdots + \lambda^{n-1} \alpha_{i_1}^{(1)} \cdots \alpha_{i_n}^{(n)}, \tag{4.2}$$

where $\wp_j$ denotes the collection of j-element-subsets of $\{1, 2, \cdots, n\}$, and $\alpha^{(1)}$, $\alpha^{(2)}, \cdots$ are used instead of $\alpha, \beta, \gamma, \cdots$, $1 \le i_t \le l_t$ for $t = 1, 2, \cdots, n$.

Proposition 4.2. The interaction plot $y_{i_1 i_2 \cdots i_n}$ *vs* $\alpha_{i_t}^{(t)}$ of the model (4.2) for n-way data of dimension $l_1 \times l_2 \times \cdots \times l_n$ consists of $\frac{l_1 l_2 \cdots l_n}{l_t}$ straight lines all intersecting at the point $(-\frac{1}{\lambda}, \mu - \frac{1}{\lambda})$.

Proof: A straightforward computation reveals that the expression above can be written as

$$y_{i_1 i_2 \cdots i_n} = \mu - \frac{1}{\lambda} + \lambda^{n-1} (\alpha_{i_1}^{(1)} + \frac{1}{\lambda}) (\alpha_{i_2}^{(2)} + \frac{1}{\lambda}) \cdots (\alpha_{i_n}^{(n)} + \frac{1}{\lambda}) \tag{4.2a}$$

and thus the conclusion holds. □

As a feasible application of this graphical technique, one can study the plot in a similar manner to detect special form of interactions for three-way data array. For instance, in the three-way ANOVA model (3.1) or (3.2), if the interaction plot y_{ijk} *vs* β_j consists of straight lines which are all parallel to each other, one can conclude that there is no interaction between factor B and the combined factor A∗C and thus the model may be reduced to

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\gamma)_{ik} \tag{4.3}$$

which is the usual three-factor randomized block design taking β_j as the jth block effect, as mentioned in the remark at the end of Chapter 3.

Suppose further that the following model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \lambda \alpha_i^* \gamma_k^* \tag{4.4}$$

is suspected and the interaction plot y_{ijk} *vs* γ_k consists of *bundle of straight lines.* Then one can speculate some kind of linearity between γ_k and γ_k^*, so the model becomes

$$y_{ijk} = \mu + \alpha_i + \beta_j + (1 + \lambda\alpha_i^*)\,\gamma_k \tag{4.5}$$

which can be also investigated.

Using the various interaction plots to detect the presence of special types of interactions in three-way array or even higher dimensional data seems to be straightforward. Even though the use of the interaction plot appears to be limited to such special types as nullity, concurrence, and particular linearity between the variables, it may be extended to any known function of the main effect parameters to model the interaction.

The next two sections use biplots and joint plots to model the interaction terms; the method uses multiplicative interaction models which do not require the experimenter to model the interaction in terms of main-effect parameters as required for those in this section.

4.3 MODEL DIAGNOSIS USING BIPLOTS FOR TWO-WAY TABLES

The purpose of this section is to review biplot methods and to provide background for the graphical procedures.

Gabriel (1971) originally introduced the two-dimensional biplot for a data matrix of rank

two as a graphical method for displaying multiple effects in a single diagram. The "bi" of biplot refers to the two-modes not two-dimensions (although higher dimensions are possible). The result below of Bradu and Gabriel (1978) is a useful form of the two-dimensional biplot for two-way data.

Suppose the rank of an $l \times m$ data matrix Y is s. Then by the singular value decomposition, Y can be decomposed as

$$Y = U\Lambda V' = (U\Lambda^{\alpha})\,(V\Lambda^{1-\alpha})' = GH', \tag{4.6}$$

where,

$$G = U\Lambda^{\alpha} = \begin{bmatrix} \underline{g}_1' \\ \underline{g}_2' \\ \vdots \\ \underline{g}_l' \end{bmatrix} \quad \text{and} \quad H = V\Lambda^{1-\alpha} = \begin{bmatrix} \underline{h}_1' \\ \underline{h}_2' \\ \vdots \\ \underline{h}_m' \end{bmatrix} \quad \text{for some } 0 \le \alpha \le 1.$$

With this decomposition, the matrix Y is represented in an s dimensional vector space by the $l+m$ vectors $\underline{g}_i'$ and $\underline{h}_j'$ by means of the usual inner product, namely $y_{ij} = \underline{g}_i' \underline{h}_j$ where $\underline{g}_i'$ and $\underline{h}_j'$ are the ith row of $G_{l \times s}$ and jth row of $H_{m \times s}$, respectively. The biplot is a plot of points on the s-dimensional Euclidean space which represents Y through the row vectors of G and H. It is usually hard to visualize such an interpretation when $s \ge 3$, while it can be fully displayed in the two-dimensional plane if $s = 2$.

When the rank s is greater than two, one may approximate y_{ij} by the inner product of two dimensional vectors with a certain level of fit. Namely, the approximation of Y by the first two components of the SVD

$$Y = [\underline{u}_1 \ , \ \underline{u}_2] \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \underline{v}_1' \\ \underline{v}_2' \end{bmatrix} = GH' \tag{4.6a}$$

where

$$G = [\underline{u}_1 , \underline{u}_2] \begin{bmatrix} \theta_1 & 0 \\ 0 & \theta_2 \end{bmatrix}^{\alpha}, \qquad H = [\underline{v}_1 , \underline{v}_2] \begin{bmatrix} \theta_1 & 0 \\ 0 & \theta_2 \end{bmatrix}^{1-\alpha}$$

achieves the highest goodness of fit of all rank 2 approximation to Y among all possible decompositions of Y by two matrices G and H of order l x 2 and m x 2. A plot of points on the plane which represents the two dimensional row vectors $\underline{g}_i'$ of G and $\underline{h}_j'$ of H is called a ***biplot.*** The biplot has been used in various contexts as an aid in performing data analysis most notably by Gabriel (1981).

Bradu and Gabriel (1978) considered some models for a two-way layout and determined the relationship between those models and the geometry of a biplot of a rank two approximation to the data matrix. Their main results may be summarized as follows:

Theorem (Bradu and Gabriel).

1. Y is closely fitted by the additive model, $y_{ij} = \mu + \alpha_i + \beta_j$, if and only if the row vectors of G are nearly collinear and the row vectors of H are nearly collinear and the two lines are nearly orthogonal.

2. Y is closely fitted by Tukey's (1949) concurrent model, $y_{ij} = \mu + \alpha_i + \beta_j + \lambda\alpha_i\beta_j$, if and only if the row vectors of G are nearly collinear and the row vectors of H are nearly collinear.

3. Y is closely fitted by Mandel's (1969) rows linear model, $y_{ij} = \mu + \alpha_i + \beta_j + \lambda\alpha_i^*\beta_j$, if and only if the row vectors of H are nearly collinear.

4. Y is closely fitted by Mandel's columns linear model, $y_{ij} = \mu + \alpha_i + \beta_j + \lambda\alpha_i\beta_j^*$, if and only if the row vectors of G are nearly collinear.

In the above diagnostic method if the rank of Y is not exactly two, one needs the rank two approximation of Y. In such a case the fit of the rank two approximation must be justified.

Gower (1990) pointed out that the biplot for the more complex model

$$y_{ij} = \mu + \alpha_i + \beta_j + \alpha_i^* \beta_j^* \quad \text{or} \quad Y = GH' = \left[\mu \underline{1}_l + \underline{\alpha},\ \underline{1}_l,\ \underline{\alpha}^* \right] \begin{bmatrix} \underline{1}'_m \\ \underline{\beta}' \\ \underline{\beta}^{*\prime} \end{bmatrix} \tag{4.7}$$

must be three-dimensional as the points given by the rows of G and H are each coplanar on two orthogonal planes. He discussed the three-dimensional geometry of biplots for more general rank 3 models of a two-way data matrix Y which fits the model

$$Y = \mu \underline{1}_l \underline{1}'_m + \underline{\alpha}\, \underline{1}'_m + \underline{1}_l\, \underline{\beta}' + \underline{\alpha}^* \underline{\beta}^{*\prime} + \underline{\alpha}^{**} \underline{\beta}^{**\prime} \tag{4.8}$$

where $\underline{\alpha}$ and/or $\underline{\beta}$ are linear combinations of $\underline{\alpha}^*$ and $\underline{\alpha}^{**}$, and $\underline{\beta}^*$ and $\underline{\beta}^{**}$, respectively.

The following basic lemmas will be used in the sequel. These lemmas are essentially due to Bradu and Gabriel who used them to prove their biplot geometry for the dimension two case. However, in the spirit of Gower, we will state them for an arbitrary rank s.

For a given $l \times m$ matrix A, let V(A) denote the vector space spanned by the column vectors, $\underline{a}_{(1)}, \underline{a}_{(2)}, \cdots, \underline{a}_{(m)}$ of A.

Lemma 4.3.1. Suppose F, G, and H are $l \times m$, $l \times s$, and $m \times s$ matrices of rank s, respectively, and $F = GH'$. Then (1) $V(F) = V(G)$ and (2) $V(F') = V(H')$.

Proof: Since $F = GH'$, one can write

$$\underline{f}_{(j)} = h_{j1}\, \underline{g}_{(1)} + h_{j2}\, \underline{g}_{(2)} + \cdots + h_{js}\, \underline{g}_{(s)} ,$$

and thus $\underline{f}_{(j)} \in V(G)$ for all j. Hence $V(F) \subseteq V(G)$. But $V(F) = V(G)$ since their dimensions are the same. This completes the proof for (1). The proof for (2) is the same. □

Lemma 4.3.2. Let G be an $l \times s$ matrix of rank s. Then all the points $(g_{11}, g_{12}, \cdots, g_{1s})$, $(g_{21}, g_{22} \cdots, g_{2s})$, $\cdots$, $(g_{l1}, g_{l2} \cdots, g_{ls})$ lie on the same hyperplane in the s-dimensional Euclidean space if and only if $\underline{1}_l \in V(G)$.

Proof: Suppose the l points $(g_{i1}, g_{i2}, \cdots, g_{is})$ lie on the same hyperplane. Then $\varphi_1 g_{i1} + \varphi_2 g_{i2} + \cdots + \varphi_s g_{is} = 1$ for some $\varphi_1, \varphi_2, \cdots, \varphi_s$ and for all $i = 1, 2, \cdots, l$. That is, $\varphi_1 \underline{g}_{(1)} + \varphi_2 \underline{g}_{(2)}, + \cdots + \varphi_s \underline{g}_{(s)} = \underline{1}_l$ and thus $\underline{1}_l \in V(G)$.

Conversely, if $\underline{1}_l \in V(G) = < \underline{g}_{(1)}, \underline{g}_{(2)}, \cdots, \underline{g}_{(s)} >$, then there exist $\rho_1, \rho_2, \cdots, \rho_s$ such that $\underline{1}_l = \rho_1 \underline{g}_{(1)} + \rho_2 \underline{g}_{(2)} + \cdots + \rho_s \underline{g}_{(s)}$. Thus $\rho_1 g_{i1} + \rho_2 g_{i2} + \cdots + \rho_s g_{is} = 1$ for all $i = 1, 2, \cdots, l$, and all points lie on the hyperplane $\rho_1 x_1 + \rho_2 x_2 + \cdots + \rho_s x_s = 1$. This completes the proof□

Notice that the hyperplane in 2-dimensional Euclidean space is a line and it is a plane in 3-dimensional space. Thus the lemma states that when $s = 2$, the points (g_{i1}, g_{i2}) are collinear if and only if $\underline{1}_l \in V(G)$ and when $s = 3$, the points (g_{i1}, g_{i2}, g_{i3}) are coplanar if and only if $\underline{1}_l \in V(G)$.

Corollary 4.3.1. For any decomposition $F = GH'$ of rank s, the points $(g_{i1}, g_{i2}, \cdots, g_{is})$ lie on the same hyperplane in s-dimensional Euclidean space if and only if $\underline{1}_l \in V(F)$.

Proof: It is an immediate consequences of the above two lemmas. □

Next, the use of joint plots for diagnosing the forms of various interactions between the variables and the type of model to fit three-way data will be discussed.

4.4 MODEL DIAGNOSIS USING JOINT PLOTS FOR THREE-WAY TABLES

Although not as simple as the biplot model, the three-mode principal component model (2.8) is also useful for representing three-way tables of data. Below it is shown that the Tucker3 model is useful as a tool for assessing appropriate models for three-way data. In particular, we show that many special forms of the general model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} \tag{4.9}$$

can be distinguished.

4.4.1 Definition of joint plots

First, the definition of the joint plots will be given. Let Y be a $l \times m \times n$ three-way data and C be the $s \times t \times u$ core array of Y. Note here that in contrast to Chapter 3, the data array Y is decomposed not the residual array D. The Tucker3 model suggests data Y be decomposed as (2.9) with each of the frontal(F_k), horizontal(H_i), and lateral(L_j) slices partitioned by the bistructure expressions. Namely, for $k = 1, \cdots, n$,

$$F_k = \begin{bmatrix} y_{11k} & \cdots & y_{1mk} \\ \vdots & & \vdots \\ y_{l1k} & \cdots & y_{lmk} \end{bmatrix} = \sum_{r=1}^{u} e_{kr} G C_r H' = \sum_{r=1}^{u} e_{kr} G^{(r)} H^{(r)'}, \tag{4.10}$$

where $G^{(r)} = GU_r\Lambda_r^{\alpha}$ and $H^{(r)} = HV_r\Lambda_r^{(1-\alpha)}$ with $C_r = U_r\Lambda_r V_r'$, the singular value decomposition of the rth frontal slice C_r of C. Similarly, for H_i and L_j, we have, for $i = 1, \cdots, l$,

$$H_i = \begin{bmatrix} y_{i11} & \cdots & y_{i1n} \\ \vdots & & \vdots \\ y_{im1} & \cdots & y_{imn} \end{bmatrix} = \sum_{p=1}^{s} g_{ip} H C_p E' = \sum_{p=1}^{s} g_{ip} H^{(p)} E^{(p)'} \tag{4.11}$$

and, for $j = 1, \cdots, m$,

$$L_j = \begin{bmatrix} y_{1j1} & \cdots & y_{1jn} \\ \vdots & & \vdots \\ y_{lj1} & \cdots & y_{ljn} \end{bmatrix} = \sum_{q=1}^{t} h_{jq} G C_q E' = \sum_{q=1}^{t} h_{jq} G^{(q)} E^{(q)'} \tag{4.12}$$

where C_p and C_q are the pth horizontal slice and the qth lateral slice of the core array C, and $H^{(p)}E^{(p)'}$ and $G^{(q)}E^{(q)'}$ are the corresponding bistructure expressions of the matrices HC_pE' and GC_qE', respectively.

The rth joint plot with respect to G and H, which will be called the GH-joint plot or GH-plot, displays the points from the rows of $G^{(r)}$ and those of $H^{(r)}$ in (4.10) on the s-dimensional Euclidean space, both $G^{(r)}$ and $H^{(r)}$ necessarily of rank $s\,(=t)$. The pth HE-joint plot on $t\,(=u)$-dimensional space and the qth GE-joint plot on $s\,(=u)$-dimensional space are defined accordingly.

Suppose distinct s, t, and u are proposed for a data array Y in the alternating least squares approximation. Then the joint plots for Y may be obtained only after adjusting the length of the pairs, G and H, H and E, and G and E. It is our custom that if $s \leq t$, then in order to obtain the s-dimensional GH-joint plot, we delete the last $t-s$ columns for H as if we only take the first s principal components for the factor correspond to H. In this case, a goodness of fit for the data, which is obtained from the joint plot

representation, is also changed due to the alteration. Thus for some three-way data array, one needs to study many joint plots simultaneously, for instance, $3s$ joint plots all together if $s = t = u \geq 2$, to diagnose the model which fits data, while for a two-way data matrix only one biplot is needed.

4.4.2. Models

A number of specializations of the general model (4.9) may be considered in practice. Among others, the following forms of special models (4.13)-(4.22) and their variations are of particular interest. Here, as in Bradu and Gabriel (1978), we name each model rather than simply list it. The names starting with "F-", "H-", and "L-" are used to explain phenomenon occurring within the frontals, horizontals, and laterals, respectively. On the other hand, "F^*-", putting $*$ on the shoulder, is used if the phenomenon happens not inside of each frontal but for interrelations between the frontals. It will be applied similarly for "H^*-" and "L^*-". Notational convention: In what follows, the superscripted Greek alphabets α^a, α^b, α^c, $\cdots$ will be used instead of α^*, α^{**}, α^{***}, $\cdots$, for simplicity.

(a) A simple standard model (the most general case)

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \lambda\alpha_i^a\beta_j^a + \theta\alpha_i^b\gamma_k^b + \nu\beta_j^c\gamma_k^c + \eta\alpha_i^d\beta_j^d\gamma_k^d \tag{4.13}$$

(b) F-bilinear model: $(\alpha_i^x \propto \alpha_i,\ \beta_j^x \propto \beta_j)$

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \lambda\alpha_i\beta_j + \theta\alpha_i\gamma_k^b + \nu\beta_j\gamma_k^c + \eta\alpha_i\beta_j\gamma_k^d \tag{4.14}$$

or equivalently,

$$F_k = \left[\underline{1}_l\ ,\ \underline{\alpha}\ \right] \begin{bmatrix} \mu+\gamma_k & 1+\nu\gamma_k^c \\ 1+\theta\gamma_k^b & \lambda+\eta\gamma_k^d \end{bmatrix} \begin{bmatrix} \underline{1}'_m \\ \underline{\beta}' \end{bmatrix}$$

$$H_i = \left[\underline{1}_m \, , \, \underline{\beta}\right] \begin{bmatrix} (\mu+\alpha_i)\underline{1}'_n + \underline{\gamma}' + \theta\alpha_i\underline{\gamma}^{b\prime} \\ (1+\lambda\alpha_i)\underline{1}'_n + \nu\underline{\gamma}^{c\prime} + \eta\alpha_i\underline{\gamma}^{d\prime} \end{bmatrix}$$

$$L_j = \left[\underline{1}_l \, , \, \underline{\alpha}\right] \begin{bmatrix} (\mu+\beta_j)\underline{1}'_n + \underline{\gamma}' + \nu\beta_j\underline{\gamma}^{c\prime} \\ (1+\lambda\beta_j)\underline{1}'_n + \theta\underline{\gamma}^{b\prime} + \eta\beta_j\underline{\gamma}^{d\prime} \end{bmatrix}$$

(c) H-bilinear model: $(\beta_j^x \propto \beta_j,\ \gamma_k^x \propto \gamma_k)$

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \lambda\alpha_i^a\beta_j + \theta\alpha_i^b\gamma_k + \nu\beta_j\gamma_k + \eta\alpha_i^d\beta_j\gamma_k \tag{4.15}$$

or equivalently,

$$F_k = \left[\, (\mu+\gamma_k)\underline{1}_l + \underline{\alpha} + \theta\gamma_k\,\underline{\alpha}^b \, , \, (1+\nu\gamma_k)\underline{1}_l + \lambda\underline{\alpha}^a + \eta\gamma_k\underline{\alpha}^d \,\right] \begin{bmatrix} \underline{1}'_m \\ \underline{\beta}' \end{bmatrix},$$

$$H_i = \left[\underline{1}_m \, , \, \underline{\beta}\right] \begin{bmatrix} \mu+\alpha_i & 1+\theta\alpha_i^b \\ 1+\lambda\alpha_i^a & \nu+\eta\alpha_i^d \end{bmatrix} \begin{bmatrix} \underline{1}'_n \\ \underline{\gamma}' \end{bmatrix}$$

(d) L-bilinear model: $(\alpha_i^x \propto \alpha_i,\ \gamma_k^x \propto \gamma_k)$

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \lambda\alpha_i\beta_j^a + \theta\alpha_i\gamma_k + \nu\beta_j^c\gamma_k + \eta\alpha_i\beta_j^d\gamma_k \tag{4.16}$$

or equivalently,

$$F_k = \left[\underline{1}_l \, , \, \underline{\alpha}\right] \begin{bmatrix} (\mu+\gamma_k)\underline{1}'_m + \underline{\beta}' + \nu\gamma_k\underline{\beta}^{c\prime} \\ (1+\theta\gamma_k)\underline{1}'_m + \lambda\underline{\beta}^{a\prime} + \eta\gamma_k\underline{\beta}^{d\prime} \end{bmatrix}$$

$$L_j = \left[\underline{1}_l \, , \, \underline{\alpha}\right] \begin{bmatrix} \mu+\beta_j & 1+\nu\beta_j^c \\ 1+\lambda\beta_j^a & 1+\eta\beta_j^d \end{bmatrix} \begin{bmatrix} \underline{1}_n' \\ \underline{\gamma}' \end{bmatrix}$$

(e) F*-linear model: $(\gamma_k^x \propto \gamma_k)$

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + \theta\alpha_i^b\gamma_k + \nu\beta_j^c\gamma_k + (\alpha\beta)_{ij}^*\gamma_k \tag{4.17}$$

that is,

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \left(1 + \theta\alpha_i^b + \nu\beta_j^c + (\alpha\beta)_{ij}^*\right)\gamma_k$$

or equivalently,

$$H_i = \left[(\mu+\alpha_i)\underline{1}_m + \underline{\beta} + (\underline{\alpha\beta})_i \, , \; (1+\theta\alpha_i^b)\underline{1}_m + \nu\underline{\beta}^c + (\underline{\alpha\beta})_i^* \right] \begin{bmatrix} \underline{1}_n' \\ \underline{\gamma}' \end{bmatrix}$$

$$L_j = \left[(\mu+\beta_j)\underline{1}_l + \underline{\alpha} + (\underline{\alpha\beta})_j \, , \; (1+\nu\underline{\beta}_j^c)\,\underline{1}_l + \theta\underline{\alpha}^b + (\underline{\alpha\beta})_j^* \right] \begin{bmatrix} \underline{1}_n' \\ \underline{\gamma}' \end{bmatrix}$$

(f) H*-linear model: $(\alpha_i^x \propto \alpha_i)$

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \lambda\alpha_i\beta_j^a + \theta\alpha_i\gamma_k^b + (\beta\gamma)_{jk} + (\beta\gamma)_{jk}^*\,\alpha_i \tag{4.18}$$

or equivalently,

$$F_k = \left[\underline{1}_l \, , \, \underline{\alpha}\right] \begin{bmatrix} (\mu+\gamma_k)\underline{1}_m' + \underline{\beta}' + (\underline{\beta\gamma})_k' \\ (1+\theta\gamma_k^b)\underline{1}_m' + \lambda\underline{\beta}^{a\prime} + (\underline{\beta\gamma})_k^{*\prime} \end{bmatrix}$$

(g) L*-linear model: $(\beta_j^x \propto \beta_j)$

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \lambda\alpha_i^a\beta_j + (\alpha\gamma)_{ik} + \nu\beta_j\gamma_k^c + (\alpha\gamma)_{ik}^* \beta_j \tag{4.19}$$

or equivalently,

$$F_k = \left[(\mu+\gamma_k)\underline{1}_l + \underline{\alpha} + (\underline{\alpha\gamma})_k \ , \ (1+\nu\gamma_k^c)\underline{1}_l + \lambda\underline{\alpha}^a + (\underline{\alpha\gamma})_k^* \right] \begin{bmatrix} \underline{1}'_m \\ \underline{\beta}' \end{bmatrix}$$

(h) *F*-interaction-free model:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} \tag{4.20}$$

(i) Trilinear model:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \lambda\alpha_i\beta_j + \theta\alpha_i\gamma_k + \nu\beta_j\gamma_k + \eta\alpha_i\beta_j\gamma_k \tag{4.21}$$

or equivalently,

$$F_k = \left[\underline{1}_l \ , \ \underline{\alpha}\right] \begin{bmatrix} \mu+\gamma_k & 1+\nu\gamma_k \\ 1+\theta\gamma_k & \lambda+\eta\gamma_k \end{bmatrix} \begin{bmatrix} \underline{1}'_m \\ \underline{\beta}' \end{bmatrix}$$

$$H_i = \left[\underline{1}_m \ , \ \underline{\beta}\right] \begin{bmatrix} \mu+\alpha_i & 1+\theta\alpha_i \\ 1+\lambda\alpha_i & \nu+\eta\alpha_i \end{bmatrix} \begin{bmatrix} \underline{1}'_n \\ \underline{\gamma}' \end{bmatrix}$$

$$L_j = \left[\underline{1}_l \ , \ \underline{\alpha}\right] \begin{bmatrix} \mu+\beta_j & 1+\nu\beta_j \\ 1+\lambda\beta_j & \theta+\nu\beta_j \end{bmatrix} \begin{bmatrix} \underline{1}'_n \\ \underline{\gamma}' \end{bmatrix}$$

(j) Additive model:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k \tag{4.22}$$

or equivalently,

$$F_k = [\underline{1}_l , \underline{\alpha}] \begin{bmatrix} \mu + \gamma_k & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \underline{1}'_m \\ \underline{\beta}' \end{bmatrix}$$

$$H_i = [\underline{1}_m , \underline{\beta}] \begin{bmatrix} \mu + \alpha_i & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \underline{1}'_n \\ \underline{\gamma}' \end{bmatrix}$$

$$L_j = [\underline{1}_l , \underline{\alpha}] \begin{bmatrix} \mu + \beta_j & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \underline{1}'_n \\ \underline{\gamma}' \end{bmatrix}$$

It will be convenient to refer to the data array Y for which a certain model fits by the name of the model. Thus, for example, Y will be said to be "F-bilinear" if (4.14) holds, etc. Beside the above listed models, there will be many other special models of (4.9). They will be named in the same manner following the rule given at the beginning of the current subsection, perhaps except for the additive model: $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k$ which is identical to the (4.21) with the condition that $\lambda = \theta = \nu = \eta = 0$.

The main purpose of studying the joint plots of three-way tables is to select a model which fits the data or parts of it. However, detailed inspection of the bistructure representation given by (4.10) to (4.12) can also give some idea about the parameters of these models. So we shall see how the estimated parameters $\hat{\mu}$, $\hat{\alpha}_i$, $\hat{\beta}_j$, $(\hat{\alpha\beta})_{ij}$, etc., can be represented by means of the inner products before we investigate the geometry of the joint plots.

4.4.3. Conversion formulas for the fitted parameters

Let Y be an $l \times m \times n$ data array, and let C be the $s \times t \times u$ core array of Y which gives the best (it is assumed here that there is an exact fit) least-squares fit $Y = GC(E \otimes H)'$ as in Section 2.2. Thus the typical element y_{ijk} of Y is best approximated by (2.8) or equivalently, the kth frontal slice of Y is given by (4.10).

As was discussed in (4.10), if $s = t$, then the GH-joint plot uses the bistructure expression for each slice as

$$F_k = \sum_{r=1}^{u} e_{kr} G^{(r)} H^{(r)'}, \qquad k = 1, \cdots, n$$

with $GC_rH' = G^{(r)}H^{(r)'}$, or equivalently

$$y_{ijk} = \sum_{r=1}^{u} e_{kr}\, \underline{g}_i^{(r)'}\, \underline{h}_j^{(r)} \qquad \text{for every } i, j, k \tag{4.23}$$

where the s-dimensional column vectors $\underline{g}_i^{(r)}$ and $\underline{h}_j^{(r)}$ are coming from the ith row of $G^{(r)}$ and the jth row of $H^{(r)}$, respectively.

Using the alternating least squares approximation for Y (cf. Section 2.2.3), the fitted parameters of the general model (4.9) for Y can be expressed in terms of the inner products of vectors in (4.23) and are given in Table 15. In the table, the following notations are used:

$$\underline{g}_.^{(r)} = \frac{1}{l} \sum_{i=1}^{l} \underline{g}_i^{(r)}, \quad \underline{h}_.^{(r)} = \frac{1}{m} \sum_{j=1}^{m} \underline{h}_j^{(r)}, \quad \text{and} \quad e_{.r} = \frac{1}{n} \sum_{k=1}^{n} e_{kr},$$

so that, for instance,

$$\bar{y}_{i..} = \sum_{r=1}^{u} e_{.r}\, \underline{g}_i^{(r)'}\, \underline{h}_.^{(r)} \quad \text{and} \quad \bar{y}_{.jk} = \sum_{r=1}^{u} e_{kr}\, \underline{g}_.^{(r)'}\, \underline{h}_j^{(r)}.$$

Table 15. Fitted parameters of the decomposition based on the GH-joint plot

Fitted Parameters	Equivalent observational expression	Equivalent Bistructure expression
$\hat{\mu}$	$\bar{y}_{...}$	$\sum_{r=1}^{u} e_{.r}\, \underline{g}_{.}^{(r)'}\, \underline{h}_{.}^{(r)}$
$\hat{\alpha}_i$	$\bar{y}_{i..}-\bar{y}_{...}$	$\sum_{r=1}^{u} e_{.r}\, (\underline{g}_{i}^{(r)}-\underline{g}_{.}^{(r)})'\, \underline{h}_{.}^{(r)}$
$\hat{\beta}_j$	$\bar{y}_{.j.}-\bar{y}_{...}$	$\sum_{r=1}^{u} e_{.r}\, \underline{g}_{.}^{(r)'}(\underline{h}_{j}^{(r)}-\underline{h}_{.}^{(r)})$
$\hat{\gamma}_k$	$\bar{y}_{..k}-\bar{y}_{...}$	$\sum_{r=1}^{u} (e_{kr}-e_{.r})\, \underline{g}_{.}^{(r)'}\, \underline{h}_{.}^{(r)}$
$(\widehat{\alpha\beta})_{ij}$	$\bar{y}_{ij.}-\bar{y}_{i..}-\bar{y}_{.j.}+\bar{y}_{...}$	$\sum_{r=1}^{u} e_{.r}\, (\underline{g}_{i}^{(r)}-\underline{g}_{.}^{(r)})'(\underline{h}_{j}^{(r)}-\underline{h}_{.}^{(r)})$
$(\widehat{\alpha\gamma})_{ik}$	$\bar{y}_{i.k}-\bar{y}_{i..}-\bar{y}_{..k}+\bar{y}_{...}$	$\sum_{r=1}^{u} (e_{kr}-e_{.r})\, (\underline{g}_{i}^{(r)}-\underline{g}_{.}^{(r)})'\, \underline{h}_{.}^{(r)}$
$(\widehat{\beta\gamma})_{jk}$	$\bar{y}_{.jk}-\bar{y}_{.j.}-\bar{y}_{..k}+\bar{y}_{...}$	$\sum_{r=1}^{u} (e_{kr}-e_{.r})\, \underline{g}_{.}^{(r)'}(\underline{h}_{j}^{(r)}-\underline{h}_{.}^{(r)})$
$(\widehat{\alpha\beta\gamma})_{ijk}$	$y_{ijk}-\bar{y}_{ij.}-\bar{y}_{i.k}-\bar{y}_{.jk}+\bar{y}_{i..}+\bar{y}_{.j.}+\bar{y}_{..k}-\bar{y}_{...}$	$\sum_{r=1}^{u} (e_{kr}-e_{.r})\, (\underline{g}_{i}^{(r)}-\underline{g}_{.}^{(r)})'(\underline{h}_{j}^{(r)}-\underline{h}_{.}^{(r)})$

As stated earlier, two other sets of conversion formulae are obtainable by using the bistructure expressions (4.11) and (4.12) for horizontal slices H_i and lateral slices L_j. One can obtain these by using the expressions

$$y_{ijk} = \sum_{p=1}^{s} g_{ip}\, \underline{h}_j^{(p)'}\, \underline{e}_k^{(p)}$$

or

$$y_{ijk} = \sum_{q=1}^{t} h_{jq}\, \underline{g}_i^{(q)'}\, \underline{e}_k^{(q)}$$

in the place of (4.23) in Table 15. However, for brevity, we omit these conversion tables as they are analogous.

4.4.4. Models and joint plots

Throughout the sequel, for any 2-dimensional joint plots of a $l \times m \times n$ data array Y, the following notations will be used: For $1 \leq r \leq u$, the rth GH-joint plot has the property L_{rG} if the l points $(g_{i1}^{(r)}, g_{i2}^{(r)})$ are collinear, and L_{rH} if the m points $(h_{j1}^{(r)}, h_{j2}^{(r)})$ are collinear. When the plot satisfies both properties and the two lines are orthogonal to each other then we will say that $L_{r\perp}$ holds. Similarly, for $1 \leq p \leq s$(resp. $1 \leq q \leq t$), we say that the pth HE-joint plot (resp. the qth GE-joint plot) has the properties L_{pH}, L_{pE}, or $L_{p\perp}$(resp. L_{qG}, L_{qE}, or $L_{q\perp}$) if they meet corresponding collinearities or orthogonality. The relationship between the geometric configurations of the joint plot and the model which may fit the data will now be investigated.

Case A. Two-dimensional joint plots:

We begin with a special case which shows how the biplot method can be generalized to a three-way data array. Suppose a $l \times m \times n$ data array Y is best approximated by a $2 \times 2 \times 1$ core. Then there is only one two-dimensional GH-joint plot available for Y.

Now the elements y_{ijk} of Y are approximated by

$$y_{ijk} = e_{k1}\, \underline{g}_i^{(1)'}\, \underline{h}_j^{(1)} \tag{4.24}$$

or equivalently the frontals are expressed by

$$F_k = e_{k1} G^{(1)} H^{(1)'}, \qquad k = 1, \cdots, n$$

where $G^{(1)}$ and $H^{(1)}$ are $l \times 2$ and $m \times 2$ matrices of rank 2. In particular, all the frontal slices of Y are generated by one single matrix $G^{(1)}H^{(1)'}$. That is, any pair of frontal slices differ by a scalar multiple. Hence, in this setting, we have:

Theorem 4.4.1. Let $Y = GC(H \otimes E)'$, where C is the matrix expression of a $2 \times 2 \times 1$ core array.

1. Y fits a special type of L-bilinear model if the GH-joint plot has the property L_{1G}.
2. Y fits a special type of H-bilinear model if the GH-joint plot has the property L_{1H}.
3. Y fits a special type of Trilinear model if the GH-joint plot has the properties L_{1G} and L_{1H}.

Proof: Let $F^{(1)} = G^{(1)}H^{(1)'}$, and $y_{ij}^{(1)} = \underline{g}_i^{(1)'}\underline{h}_j^{(1)}$. Suppose the GH-joint plot has the property L_{1G}. Then, by Bradu and Gabriel's Theorem, the matrix $F^{(1)}$ fits the model

$$y_{ij}^{(1)} = \mu^{(1)} + \alpha_i^{(1)} + \beta_j^{(1)} + \lambda^{(1)}\alpha_i^{(1)}\beta_j^{(1)*},$$

where, $\mu^{(1)} = \underline{g}_\cdot^{(1)'}\underline{h}_\cdot^{(1)}$, $\alpha_i^{(1)} = (\underline{g}_i^{(1)} - \underline{g}_\cdot^{(1)})'\, \underline{h}_\cdot^{(1)}$, $\beta_j^{(1)} = \underline{g}_\cdot^{(1)'}(\underline{h}_j^{(1)} - \underline{h}_\cdot^{(1)})$, and $\beta_j^{(1)*}$ are some constants with $\sum_j \beta_j^{(1)*} = 0$. From (4.24), y_{ijk} is approximated by $e_{k1}y_{ij}^{(1)}$, which is, in turn, expressed as

$$y_{ijk} = (e_{\cdot 1} + \epsilon_{k1})\ (\mu^{(1)} + \alpha_i^{(1)} + \beta_j^{(1)} + \lambda^{(1)}\alpha_i^{(1)}\beta_j^{(1)*}),$$

where, $\epsilon_{k1} = e_{k1} - e_{.1}$. Therefore, by using the conversion formulae as in Table 15, we have

$$e_{.1}\mu^{(1)} = \mu, \quad e_{.1}\alpha_i^{(1)} = \alpha_i, \quad e_{.1}\beta_j^{(1)} = \beta_j, \quad \epsilon_{k1}\mu^{(1)} = \gamma_k, \quad \lambda^{(1)}\beta_j^{(1)^*} = \lambda\beta_j^*, \quad \mu^{(1)}e_{.1} = \frac{1}{\theta}, \quad \lambda\theta = \eta,$$

and

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \lambda\alpha_i\beta_j^* + \theta\alpha_i\gamma_k + \theta\beta_j\gamma_k + \eta\alpha_i\beta_j^*\gamma_k,$$

which is L-bilinear model with $\nu\beta_j^c = \theta\beta_j$ and $\beta_j^a = \beta_j^d = \beta_j^*$. By the same argument, it is proved that if the plot has the property L_{1H}, then Y fits

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \lambda\alpha_i^*\beta_j + \nu\alpha_i\gamma_k + \nu\beta_j\gamma_k + \eta\alpha_i^*\beta_j\gamma_k.$$

Also if the plot has both properties L_{1G} and L_{1H}, then Y fits the trilinear model with $\theta = \nu$. □

We note that, from Corollary 4.3.1 and from the matrix expressions of F_k in (4.15), (4.16), and (4.21), it can be shown that the converse is also true under our assumption.

Theorem 4.4.2. If for every k, $F_k = e_{k1}GCH' = e_{k1}G^{(1)}H^{(1)'}$, where C is the matrix expression of a $2 \times 2 \times 1$ core array, then Y fits the F-interaction-free model if and only if the GH-joint plot has the property $L_{1\perp}$.

Proof: Suppose Y fits the F-interaction-free model. Then

$$y_{ijk} - y_{ij'k} - y_{i'jk} + y_{i'j'k} = 0 \quad \text{for all } i, i' \text{ and } j, j'$$

for every k. By using (4.24), this can be expressed as

$$e_{k1}\left(\underline{g}_i^{(1)'}\underline{h}_j^{(1)} - \underline{g}_i^{(1)'}\underline{h}_{j'}^{(1)} - \underline{g}_{i'}^{(1)'}\underline{h}_j^{(1)} + \underline{g}_{i'}^{(1)'}\underline{h}_{j'}^{(1)}\right) = 0.$$

This is equivalent to

$$(\underline{g}_i^{(1)} - \underline{g}_{i'}^{(1)})'(\underline{h}_j^{(1)} - \underline{h}_{j'}^{(1)}) = 0 \quad \text{for all } i, i', j, j'.$$

Thus the points $(g_{i1}^{(1)}, g_{i2}^{(1)})$ form a line and $(h_{j1}^{(1)}, h_{j2}^{(1)})$ form a line and the two lines are orthogonal to each other. Conversely suppose the joint plot has the property $L_{1\perp}$. Then by the Bradu and Gabriel's Theorem, $F^{(1)}$ fits the model

$$y_{ij}^{(1)} = \mu^{(1)} + \alpha_i^{(1)} + \beta_j^{(1)}$$

or equivalently, we can write

$$(\underline{g}_i^{(1)} - \underline{g}_{\cdot}^{(1)})'(\underline{h}_j^{(1)} - \underline{h}_{\cdot}^{(1)}) = 0 \quad \text{for all } i, j.$$

This implies that the interactions $(\alpha\widehat{\beta}\gamma)_{ijk}$ and $(\alpha\widehat{\beta})_{ij}$ in the model (4.9) are zero according to the conversion formulae given in Table 1. So Y fits the F-interaction-free-model. □

Suppose a $l \times m \times n$ data array Y is best approximated using a $2 \times 2 \times u$ core array C for some $u \geq 2$. Then there are u GH-joint plots for Y and each frontal is expressed by (4.10), or by

$$F_k = e_{k1} F^{(1)} + \cdots + e_{ku} F^{(u)}$$

with the matrices $F^{(r)} = GC_rH' = G^{(r)}H^{(r)'}$. Each joint plot may indicate a meaningful aspect about interaction and should be taken into account in order to determine a suitable model for the data. However, it is shown that if c_{111} explains most of the variation in the data, that is, by definition, the proportion of c_{111}^2 with respect to the total sum of squared observations is significantly close to 1 then the first joint plot alone summarizes most of information and $e_{k1} F^{(1)}$ itself gives a reasonably good approximation of F_k for every k. If this is the case, in many instances, the rest of the joint plots are not relevant and can be ignored. Then we will say that the frontals are controlled by the first joint plot.

In the alternating least squares algorithm, the core dimensions (s, t, u) must be specified prior to the fit. However, as indicated below, Theorem 4.4.1 and 4.4.2 are still valid even if one of the dimensions is overspecified.

Corollary 4.4.1 Let $Y = GC(E \otimes H)'$, where C is the matrix expression of a $2 \times 2 \times u$ core array for some $u \geq 2$. Suppose the frontals of Y are controlled by the first joint plot. Then the statements 1, 2, and 3 of Theorem 4.4.1 hold.

Corollary 4.4.2. Let $Y = GC(E \otimes H)'$, where C is the matrix expression of a $2 \times 2 \times u$ core array for some $u \geq 2$. Suppose either all frontals are controlled by the first GH-joint plot that has the property $L_{1\perp}$, or all the GH-joint plots have the property $L_{r\perp}$. Then Y fits the *F*-interaction-free model.

Proof: The first part of the Corollary is an immediate consequence of Theorem 4.4.2. For the second part, suppose for every r, the rth GH-joint plot has the property $L_{r\perp}$. Then using the notations in the proof of Theorem 4.4.1,

$$\begin{aligned} y_{ijk} &= \sum_{r=1}^{u} e_{kr}(\mu^{(r)} + \alpha_i^{(r)} + \beta_j^{(r)}) \\ &= \sum_{r=1}^{u} (e_{\cdot r} + \epsilon_{kr})(\mu^{(r)} + \alpha_i^{(r)} + \beta_j^{(r)}) \\ &= \sum_{r=1}^{u} e_{\cdot r}\mu^{(r)} + \sum_{r=1}^{u} e_{\cdot r}\alpha_i^{(r)} + \sum_{r=1}^{u} e_{\cdot r}\beta_j^{(r)} + \sum_{r=1}^{u} \epsilon_{kr}\mu^{(r)} + \sum_{r=1}^{u} \epsilon_{kr}\alpha_i^{(r)} + \sum_{r=1}^{u} \epsilon_{kr}\beta_j^{(r)}. \end{aligned}$$

Thus Y fits the model of the type

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk}.$$

This completes the proof. □

So far, two-dimensional GH-joint plots for data having a $2 \times 2 \times u$ core and their relationship to the suitable models which fit the data have been considered. Similarly,

analogous results are obtained by investigating the GE-joint plots for data having a $2 \times t \times 2$ core and HE-joint plots for data having a $s \times 2 \times 2$ core. In particular, if $s = t = u \geq 2$, then one needs to combine all feasible models, each of which is obtained from the inspection of GH-, GE-, and HE-joint plots, respectively.

We now summarize our main result for the case when $s = t = u = 2$ without proof.

Corollary 4.4.3. Let $Y = GC(E \otimes H)'$, where C is the matrix expression of a $2 \times 2 \times 2$ core C. Suppose the frontals, horizontals, and laterals are controlled by the first joint plots of corresponding GH-, HE-, and GE-joint plots, respectively. Then the various combinations of collinearities of the three plots and the corresponding models are given in Table 16, where the notations L_G, L_H, L_E, and $L_\perp$ are used instead of L_{1G}, L_{1H}, L_{1E}, and $L_{1\perp}$, respectively, and ξ_{ijk} indicates the type of interaction terms in the model (4.9) (see also (3.55)).

Remarks

The model for data which has a $2 \times 2 \times 2$ core and each of the second GH-, GE-, HE-joint plots has a strong effect (not as much as the first ones but still the contribution is significant) may indicate other interaction terms besides the ones already found in the first plots. Hence the model of such data has more general forms than listed in Table 16. In some of such cases, one can utilize the method that was used in the proofs of Theorem 4.4.1 and Corollary 4.4.2 to find a suitable model for the given data. For example, if all the joint plots appear as $L_\perp$, then there is no three-factor interaction. In the second GH-joint plot, if almost all $\underline{g}$-points converge to a single point and all $\underline{h}$-points also converge, then their inner products between any two vectors, one from G and one from H, are almost equal, so it does not affect the type of the model at all.

If the first joint plots of the data has poor goodness of fit, then none of the above models can be suitable and one might also try to use the joint plots in a stepwise

Table 16. Diagnostic rules for three-way tables when $s = t = u = 2$

Properties satisfied by GH-plots	GE-plots	HE-plots	Structure of ξ_{ijk} in $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \xi_{ijk}$	Type of the model for Y
$L_\perp$	$L_\perp$	$L_\perp$	0	Additive
L_G, L_H	$L_\perp$	$L_\perp$	$\lambda\alpha_i\beta_j$	F^*-interaction-free
L_G			$\lambda\alpha_i\beta_j^a$	
L_H			$\lambda\alpha_i^a\beta_j$	
$--$			$(\alpha\beta)_{ij}$	
$L_\perp$	$L_\perp$	$--$	$(\beta\gamma)_{jk}$	H^*-interaction-free
$L_\perp$	$--$	$L_\perp$	$(\alpha\gamma)_{ik}$	L^*-interaction-free
$L_\perp$	$--$	$--$	$(\alpha\gamma)_{ik} + (\beta\gamma)_{jk}$	F-interaction-free
L_G, L_H	L_G, L_E	$L_\perp$	$\lambda\alpha_i\beta_j + \theta\alpha_i\gamma_k$	H-interaction-free
L_H	L_G, L_E		$\lambda\alpha_i^a\beta_j + \theta\alpha_i\gamma_k$	
L_G, L_H	L_E		$\lambda\alpha_i\beta_j + \theta\alpha_i^b\gamma_k$	
L_H	L_E		$\lambda\alpha_i^a\beta_j + \theta\alpha_i^b\gamma_k$	
L_G	L_G		$\lambda\alpha_i\beta_j^a + \theta\alpha_i\gamma_k^b$	
$--$	L_G		$(\alpha\beta)_{ij} + \theta\alpha_i\gamma_k^b$	
L_G	$--$		$\lambda\alpha_i\beta_j^a + (\alpha\gamma)_{ik}$	
$--$	$--$		$(\alpha\beta)_{ij} + (\alpha\gamma)_{ik}$	
$--$	$L_\perp$	$--$	$(\alpha\beta)_{ij} + (\beta\gamma)_{jk}$	L-interaction-free
L_G, L_H	L_G, L_E	L_H, L_E	$\lambda\alpha_i\beta_j + \theta\alpha_i\gamma_k + \nu\beta_j\gamma_k + \eta\alpha_i\beta_j\gamma_k$	Trilinear
L_G, L_H	L_G	L_H	$\lambda\alpha_i\beta_j + \theta\alpha_i\gamma_k^b + \nu\beta_j\gamma_k^c + \eta\alpha_i\beta_j\gamma_k^d$	F-bilinear
L_H	L_E	L_H, L_E	$\lambda\alpha_i^a\beta_j + \theta\alpha_i^b\gamma_k + \nu\beta_j\gamma_k + \eta\alpha_i^d\beta_j\gamma_k$	H-bilinear
L_G	L_G, L_E	L_E	$\lambda\alpha_i\beta_j^a + \theta\alpha_i\gamma_k + \nu\beta_j^c\gamma_k + \eta\alpha_i\beta_j^d\gamma_k$	L-bilinear
$--$	L_E	L_E	$(\alpha\beta)_{ij} + \theta\alpha_i^b\gamma_k + \nu\beta_j^c\gamma_k + (\alpha\beta)_{ij}^*\gamma_k$	F^*-linear
L_G	L_G	$--$	$\lambda\alpha_i\beta_j^a + \theta\alpha_i\gamma_k^b + (\beta\gamma)_{jk} + (\beta\gamma)_{jk}^*\alpha_i$	H^*-linear
L_H	$--$	L_H	$\lambda\alpha_i^a\beta_j + (\alpha\gamma)_{ik} + \nu\beta_j\gamma_k^c + (\alpha\gamma)_{ik}^*\beta_j$	L^*-linear

manner: First, diagnose the most appropriate model through the inspection of the first joint plots, fit this, and then study the joint plots of residuals from the fit to see if this indicates a reasonable extension of the fitted model. The sum of these can serve as proper models for the data. In general, to determine a suitable model for data one has to examine the contribution and importance of each joint plot as well as the geometric configuration of the plot.

Case B. Three-dimensional joint plots:

For a given data array Y, if the approximation by a $2\times2\times2$ core seems not to suffice then it is necessary to find higher-dimensional joint plot representations. As the method is a graphical one, it is of limited appeal if the dimension is larger than three. However, for dimension 3, we may be able to extend the technique in exactly the same manner as Gower (1990) has extended the planar biplot method to the 3-dimensional biplot for the rank three data matrix. In this subsection three-dimensional joint plots will be briefly discussed for one special case to exhibit some interesting structures involved.

Consider the following fairly general model, yet another specialization of (4.9):

$$\begin{aligned} y_{ijk} = {} & \mu + \alpha_i + \beta_j + \gamma_k + (\lambda_1\alpha_i\beta_j + \lambda_2\alpha_i^a\beta_j + \lambda_3\alpha_i\beta_j^a + \lambda_4\alpha_i^a\beta_j^a) + (\theta_1\gamma_k^a\alpha_i + \theta_2\gamma_k^c\alpha_i^a) \\ & + (\nu_1\gamma_k^b\beta_j + \nu_2\gamma_k^d\beta_j^a) + (\eta_1\gamma_k^e\alpha_i\beta_j + \eta_2\gamma_k^f\alpha_i^a\beta_j + \eta_3\gamma_k^g\alpha_i\beta_j^a + \eta_4\gamma_k^h\alpha_i^a\beta_j^a) \end{aligned} \tag{4.25}$$

or equivalently,

$$F_k = \left[\underline{1}_l\ ,\ \underline{\alpha}\ ,\ \underline{\alpha}^a\right] \begin{bmatrix} \mu+\gamma_k & 1+\nu_1\gamma_k^b & \nu_2\gamma_k^d \\ 1+\theta_1\gamma_k^a & \lambda_1+\eta_1\gamma_k^e & \lambda_3+\eta_3\gamma_k^g \\ \theta_2\gamma_k^c & \lambda_2+\eta_2\gamma_k^f & \lambda_4+\eta_4\gamma_k^h \end{bmatrix} \begin{bmatrix} \underline{1}_m' \\ \underline{\beta}' \\ \underline{\beta}^{a\prime} \end{bmatrix}.$$

Notice that each interaction term in (4.9) is decomposed in terms of α_i, β_j, α_i^a, β_j^a, and γ_k^* s. This model is represented by a 3-dimensional bistructure and is subject to $\sum_i \alpha_i = \sum_j \beta_j = \sum_k \gamma_k = 0$ as well as all the sums of interaction terms are zero over i, over j, and over k.

A number of specializations of this model are recognized as models in Table 16 or their special forms. For example, if all coefficients λ, θ, ν, and η vanish then it becomes an additive model and if all except for λ_1, θ_1, ν_1, and η_1 vanish then it is the F-bilinear model. Suppose Y is a special three-way data array which is approximated by $3\times3\times1$ core array C. Then as before, the frontals of Y can be expressed by a single bistructure decomposition. That is,

$$F_k = e_{k1} G^{(1)} H^{(1)'} \quad \text{for } k = 1, \cdots, n.$$

For such data, if Y fits a particular form of the model (4.25), then the geometry of the joint plot which consists of the points given by the rows of the $l\times3$ matrix $G^{(1)}$ and the $m\times3$ matrix $H^{(1)}$ on the three-dimensional Euclidean space is characterized as follows.

Theorem 4.4.3. Let $Y = GC(E \otimes H)'$, where C is the matrix expression of a $3\times3\times1$ core array.
1. If Y fits the model (4.25) then the points given by the rows of $G^{(1)}$ and $H^{(1)}$ are each coplanar.
2. If Y fits the model (4.25) with $\lambda_1 = \lambda_2 = \lambda_3 = \eta_1 = \eta_2 = \eta_3 = 0$ then the points given by the rows of $G^{(1)}$ and $H^{(1)}$ are coplanar on two orthogonal planes.

Proof: 1 is an immediate consequence of Corollary 4.3.1. For 2, notice that if Y fits for the given model, then the model also can be decomposed as

$$F_k = e_{k1} G^* H^{*'} \text{ with}$$

$$G^* = \left[(\mu+\gamma_k)\underline{1}_l + (1+\theta_1\gamma_k^a)\underline{\alpha} + \theta_2\gamma_k^c\underline{\alpha}^a \;,\quad (1+\nu_1\gamma_k^b)\underline{1}_l \;,\quad \nu_2\gamma_k^d\underline{1}_l + (\lambda_4+\eta_4\gamma_k^h)\,\underline{\alpha}^a\right]$$

$$H^* = \left[\underline{1}_m \, , \; \underline{\beta} \, , \; \underline{\beta}^a \right].$$

This bistructure decomposition may be considered as a particular solution to the proposed model and shows that the points given by the rows of G^* and H^* are each coplanar on two orthogonal planes. We can show that the simple geometry is essentially independent of the choice of decomposition $G^*H^{*'}$ for F_k. □

It is now easy to see that all the special forms discussed in case A may arise when the F_k are decomposed as rank two bistructures, giving degenerate two-dimensional joint plots from the configuration on three-dimensional space discussed in this section. Most of the joint plots of dimension two should be interpreted as two-dimensional cross-sections of the three-dimensional joint plots .

4.5 EXAMPLE

This example illustrates the above methodology, and comes from a study of the toxicity of factory effluents. The authors, Wuhrmann and Worker (1953), measured the effects of the different concentrations of cyanide ion on the survival time of Phoxinus laevis, European minnows. Five concentrations of cyanide were combined with three concentrations of oxygen and three temperatures. For each combination of factors ten replicate fishes were examined. However, the data in Table 17 gives only the sum of the ten readings and thus do not provide an estimate of within-subgroup variance.

The usual analysis of variance table was obtained by fitting the model (3.2). Table 18 reveals that there is a significant cyanide by oxygen interaction, as well as a cyanide by

Table 17. Toxicity data

Factor3 Temperature in °C	Factor 1 Cyanide concentrations in mg/l	Factor 2 Oxygen concentrations in mg/l		
		$O_1(1.5)$	$O_2(3.0)$	$O_3(9.0)$
$T_1(5)$	$Cy_1(0.16)$	201	246	271
	$Cy_2(0.80)$	150	164	170
	$Cy_3(4.0)$	131	138	149
	$Cy_4(20.0)$	130	136	127
	$Cy_5(100.0)$	97	102	99
$T_2(15)$	$Cy_1(0.16)$	124	158	207
	$Cy_2(0.80)$	104	111	117
	$Cy_3(4.0)$	86	99	81
	$Cy_4(20.0)$	89	91	87
	$Cy_5(100.0)$	60	74	72
$T_3(25)$	$Cy_1(0.16)$	79	129	142
	$Cy_2(0.80)$	63	54	93
	$Cy_3(4.0)$	50	51	62
	$Cy_4(20.0)$	51	52	51
	$Cy_5(100.0)$	32	46	52

Note: Data by Wuhrmann and Woker (1953) (cf. Sokal and Rohlf, 1981)

Table 18. Analysis of variance table for the toxicity data

Source	df	SS	MS	F	P-value
Cyanide	4	55545.467	13886.367	214.68	0.0001
Oxygen	2	3758.800	1879.400	29.06	0.0001
Temperature	2	57116.133	28558.067	441.51	0.0001
Cyanide*Oxygen	8	5264.533	658.067	10.17	0.0001
Cyanide*Temperature	8	3685.867	460.733	7.12	0.0005
Oxygen*Temperature	4	97.067	24.267	0.38	0.8230
Error	16	1304.943	64.683		
Total	44	126502.80			

temperature interaction, but no oxygen by temperature interaction. Thus the results of the analysis of variance table indicates that the effect of the cyanide ion on survival time depends on the oxygen concentration and the temperature of the water.

The three-mode principal component analysis was applied to the data in order to illustrate how the joint plots can be used to model the interaction terms. To run the three-mode principal component analysis procedure, a SAS program (cf. chapter 6) using SAS/IML has been written following the User's guide by Kroonenberg and Brouwer (1985). First, some of the most useful information from the analysis will be discussed, i.e. component matrices, core arrays, and joint plots.

The component matrices (G, H, E) and eigenvalues (λ_p, μ_q, ν_r) are given in Table 19 with $s = t = u = 2$. The principal components, the columns of orthonormal matrices G, H, and E of a factor, are arranged in decreasing order of importance as expressed by eigenvalues. For all three factors, the first two components together account for almost all the variability in the data and the second component explains less than 1% of the total variability in the data. For illustration purpose, the following analysis will be based on a $2 \times 2 \times 2$ reduction although the $5 \times 3 \times 3$ data may be reduced quite effectively to a smaller dimension.

The $2 \times 2 \times 2$ core array in terms of two 2×2 frontal core matrices are displayed in Table 20. The core matrix represents a partitioning of the overall fitted sum of squares into small units through the relationship between the components of the three-factors. In particular, like the squared singular values in the singular value decomposition, the squared core elements c^2_{pqr} partition the fitted variation into parts which can be attributed to each component. As is often the case, especially, if the analysis is based on the non-centered data like our example, the most important element of the core array is the first element c_{111}.

In toxicity data, the high proportion of c^2_{111} (0.992) indicates that the first components of the three factors explain most of the variability in the data and this suggests that the first joint plot alone summarizes most of low and higher order effects interaction in the

Table 19. Component matrices and standardized components ($s = t = u = 2$)

Component scores of Cyanide (in mg/l)

G	g_p		$g_p \lambda_p^{1/2}$	
	$\underline{g}_1$	$\underline{g}_2$	$\underline{g}_{1\_st}$	$\underline{g}_{2\_st}$
Cyanide 1 (0.16)	0.6775	-0.6875	0.6752	-0.0474
Cyanide 2 (0.80)	0.4470	0.1742	0.4455	0.0120
Cyanide 3 (4.0)	0.3724	0.3977	0.3711	0.0274
Cyanide 4 (20.0)	0.3553	0.5643	0.3541	0.0389
Cyanide 5 (100)	0.2759	0.1424	02750	0.0098
λ_p_raw	65703.2	3118.5		
λ_p_st	0.9930	0.0048		

Component scores of Oxygen (in mg/l)

H	h_q		$h_q \mu_q^{1/2}$	
	$\underline{h}_1$	$\underline{h}_2$	$\underline{h}_{1\_st}$	$\underline{h}_{2\_st}$
Oxygen 1 (1.5)	0.5048	0.7596	0.5030	0.0535
Oxygen 2 (3.0)	0.5848	0.0485	0.5827	0.0034
Oxygen 3 (100)	0.6350	-0.6486	0.6327	-0.0456
μ_q_raw	60577.1	3244.6		
μ_q_st	0.9928	0.0050		

Component scores of Temperature (in °C)

E	e_r		$e_r \nu_r^{1/2}$	
	$\underline{e}_1$	$\underline{e}_2$	$\underline{e}_{1\_st}$	$\underline{e}_{2\_st}$
Temp 1 (5°)	0.7747	-0.5158	0.7731	-0.0226
Temp 2 (15°)	0.5270	0.2070	0.5259	0.0091
Temp 3 (25°)	0.3494	0.8313	0.3487	0.365
ν_r_raw	652561.6	1260.2		
ν_r_st	0.9959	0.0019		

Note: The standardized component weight or standardized eigenvalue λ_{p_st}, for example, indicates the relative contribution of a component to the overall fit of the model, i.e. $\sum_p \lambda_{p\_st}$ $= \mathrm{SS(Fit)} / \{\mathrm{SS(Total)}\} = \lambda_p / (\sum_p \lambda_p)$.

Table 20. Frontal planes of core array

Estimated core elements $\hat{c}_{pqr}$		Explained variation $\frac{\hat{c}^2_{pqr}}{\sum_{i=1}^{5}\sum_{j=1}^{3}\sum_{k=1}^{3} y^2_{ijk}}$		Designation of elements c_{pqr}	
Frontal plane 1					
CF_1	h_1	h_2	h_1	h_2	
g_1	806.2627	0.2446	0.9920	0.0000	c_{111} c_{121}
g_2	0.2812	50.0200	0.0000	0.0038	c_{211} c_{221}
Frontal plane 2					
CF_2	h_1	h_2	h_1	h_2	
g_1	-0.6024	-25.3642	0.0000	0.0010	c_{112} c_{122}
g_2	-22.7427	9.9611	0.0008	0.0002	c_{212} c_{222}

Note: 1. $\frac{\hat{c}^2_{pqr}}{\sum_{i=1}^{5}\sum_{j=1}^{3}\sum_{k=1}^{3} y^2_{ijk}} = \frac{\hat{c}^2_{pqr}}{\text{Uncorrected SS(Total)}}$ = is the proportion of the amount of variance jointly accounted by the pth, qth, and rth component of G, H, and E, respectively.

2. $\text{SS(fit)} = \sum_{p=1}^{2}\sum_{q=1}^{2}\sum_{r=1}^{2} \hat{c}^2_{pqr}$

data. For example, for each frontal, i.e., at each level of temperature, $||R_k||^2 = ||F_k - e_{k1} F^{(1)}||^2$ are relatively small compare to $||F_k||^2$:

For $k = 1$: $||F_1||^2 = 392199$, $||e_{11} F^{(1)}||^2 = 391641.43\,(99.8\%)$, $||R_1||^2 = 540.33\,(0.1\%)$

For $k = 2$: $||F_2||^2 = 181804$, $||e_{21} F^{(1)}||^2 = 181235.30\,(99.7\%)$, $||R_2||^2 = 562.89\,(0.3\%)$

For $k = 3$: $||F_3||^2 = 81275$, $||e_{31} F^{(1)}||^2 = 79664.96\,(98.0\%)$, $||R_3||^2 = 1613.00\,(2\%)$.

Thus the approximation of the frontals F_k, by the single term $e_{k1} F^{(1)}$, is considered as a reasonably good approximation, so one only need to focus on the first joint plot to determine an appropriate model for Y. Furthermore, if this is the case then the horizontals and the laterals are also controlled by the first joint plots of the HE-joint plot, and the GE-joint plot, respectively.

Figures 8 to 10 displays the three joint plots from the toxicity data. The most salient configuration of the first HE-joint plot in Figure 8 is that it has the properties $L_{1\perp}$, i.e., two lines on the plot show perfect collinearity and the angle between the two lines is near 89°. Thus, by Theorem 4.4.2, there is no three factor interaction and also there may be only some very negligible interaction between oxygen and temperature. These findings are consistent with the result obtained from the analysis of variance.

The first GH-joint plot in Figure 9 displays strong collinearity of the row vectors of H, but only the first four row vectors of G in the joint plot are collinear. So one may conclude that there is interaction between the cyanide and oxygen factors and the corresponding interaction term can be expressed as a linear function of β_j, i.e., a scalar multiple of the product $\alpha_i^a \beta_j$.

Similarly, Figure 10 of the first GE-joint plots indicate that the vectors of E, and the G-vectors are nearly collinear except for $\underline{g}_5$. So the GE-joint plots exhibit the properties L_{1E} and thus, by Table 16, the interaction between cyanide and temperature can be best modeled as $\alpha_i^b \gamma_k$.

Thus, by Corollary 4.4.2 and Table 16, the configurations of the joint plots suggest the special form of the H-interaction free model,

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \lambda\alpha_i^a\beta_j + \theta\alpha_i^b\gamma_k$$

for the toxicity data. Note that the coefficients λ and θ should be a rather small number as both the configurations in the joint plots and the F-ratios in the analysis of variance table indicate that the corresponding interactions appears to be not very strong.

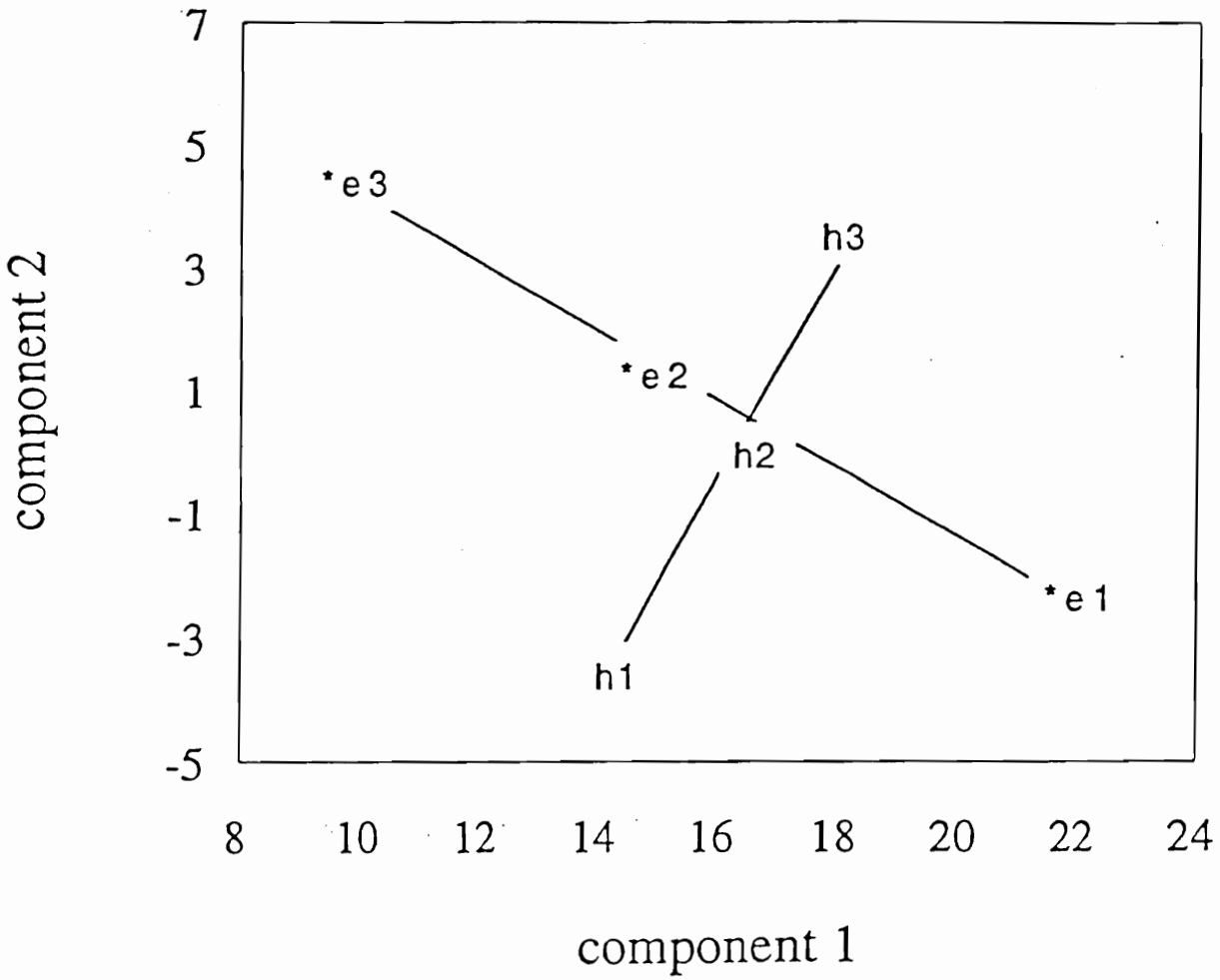


Figure 8. The HE-joint plot based on the first horizontal core matrix

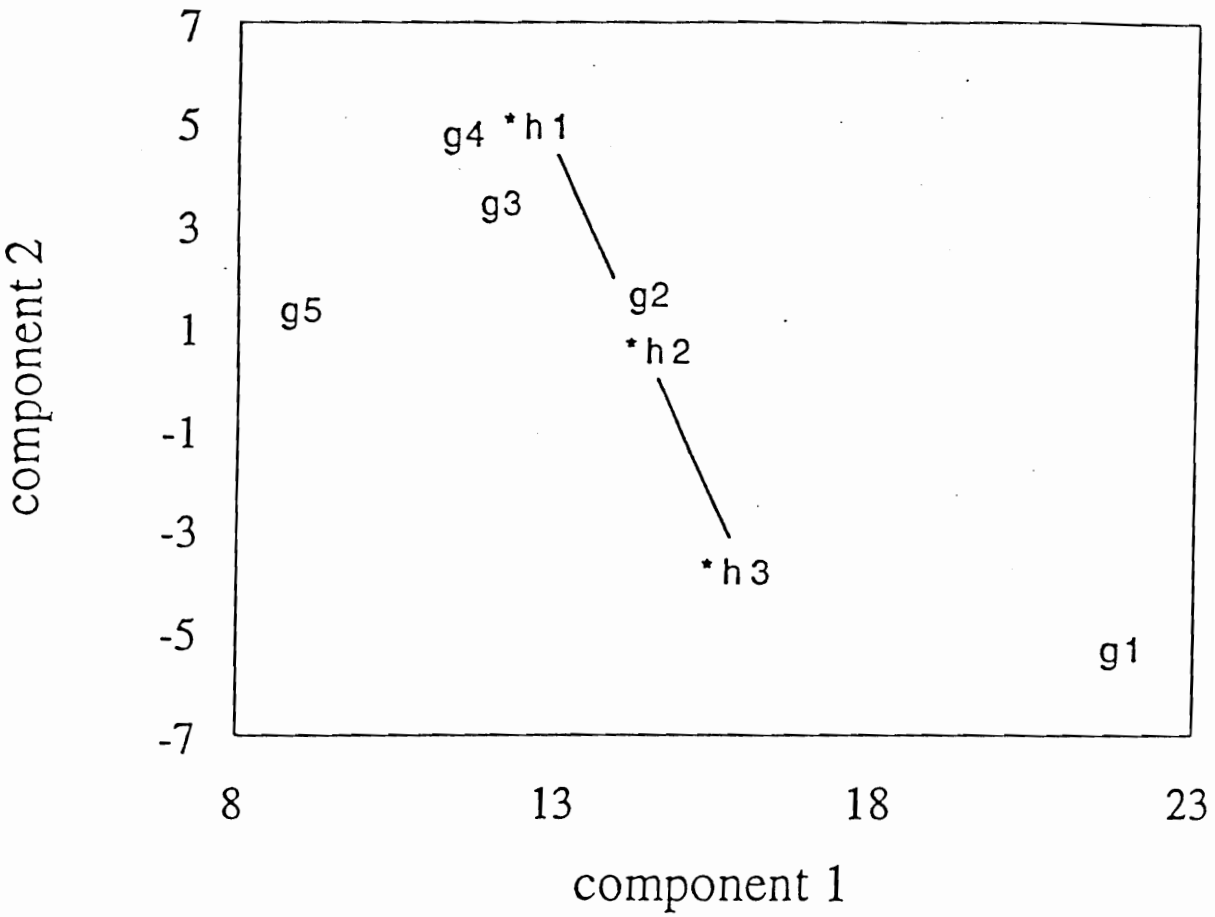


Figure 9. The GH-joint plot based on the first frontal core matrix

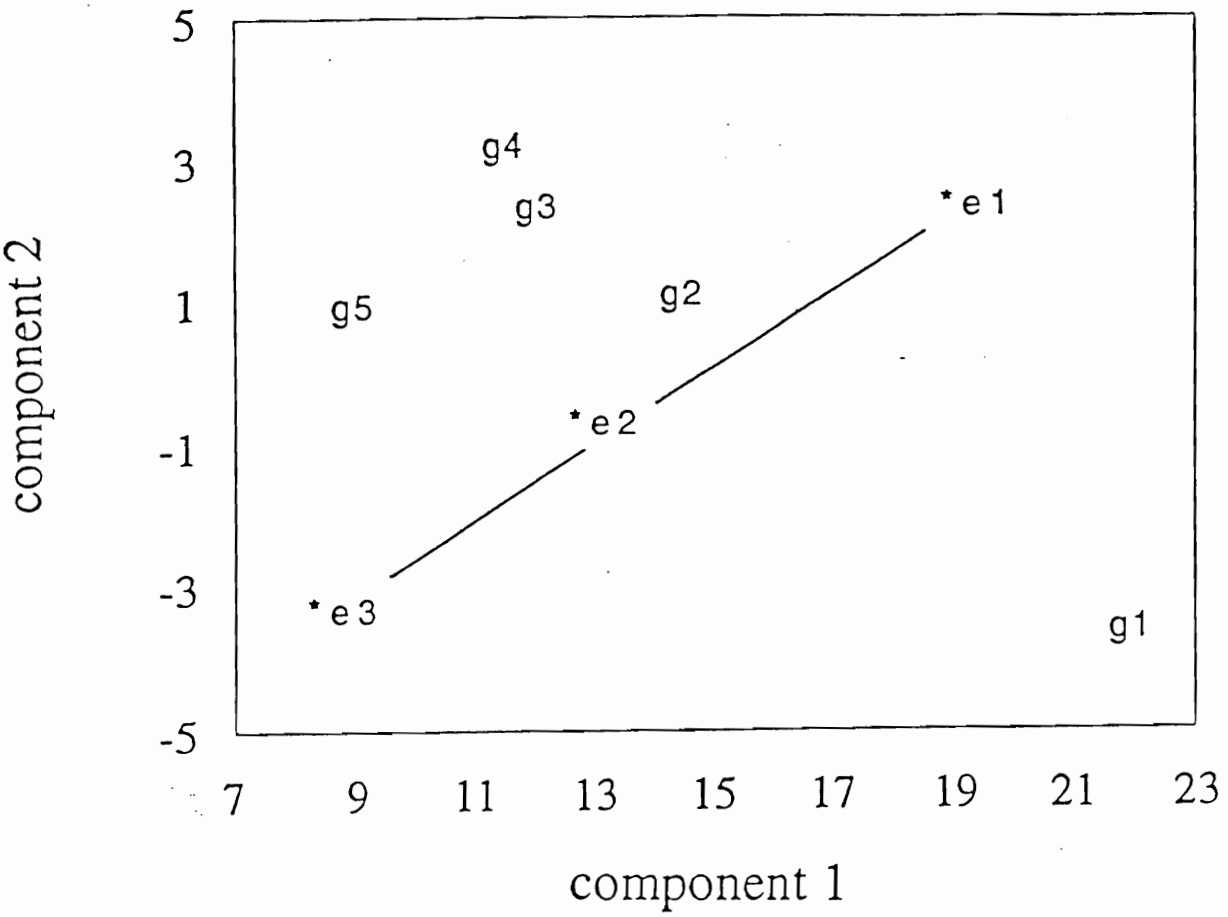


Figure 10. The GE-joint plot based on the first lateral core matrix

CHAPTER 5. CONCLUSIONS

In this chapter, the findings of this study are summarized. Also some areas of further research are addressed throughout the chapter.

The main contributions of this investigation are the development of procedures, (i) to separate three-factor interaction terms from random error, and (ii) to diagnose interaction terms using a graphical approach, in a three-way table with one observation per cell.

The dissertation is divided into two parts. First, Chapter 3 discusses testing procedures and also a simulation technique to show how to select appropriate terms for three-factor interaction. The second part in Chapter 4 investigates a graphical approach for model building. The analysis is a three-way analogue of the biplot graphical analysis for two-way tables. The dissertation applies a multivariate method, three-mode principal component analysis, to decompose three-way tables with one observation per cell.

In Chapter 3, the method is used to partition the three-factor interaction sum of squares into the portion related to the interaction and the portion related to random error. Thus the aim is to separate three-way interaction from random error by retaining the most significant interaction term in the model (3.5). The new working model is the modified standard three-way ANOVA model with multiplicative terms from the three-mode principal component analysis. Two approaches are discussed to solve the problem of testing for interaction with unreplicated data: randomization tests and likelihood ratio

tests. Moreover, a criteria to determine the number of interaction terms which will be retained in the model is proposed using Monte Carlo methods. In order to decide which way to go one needs to think about how efficient these tests are relative to one another.

For future study:
(1) There is a definite need to know how to analyse the repeated measure design using the approach we use in Chapter3. Thus, a two-way multivariate method will be studied.
(2) A table of critical values based on Monte Carlo methods is needed for different sample sizes. This requires an extensive simulation.
(3) There is a need to study for some other designs such as the Latin Squares Design which has missing observations.
(4) If a model is nonadditive and a significant interaction is observed, one may be interested in investigating the cause of the interaction, i.e. whether it is due to either a true interaction, or to factor-related variance heterogeneity, or to outliers, or to some combination of these. A general technique based on three-mode principal component analysis modeling procedure that can disentangle these types of "nonadditivity" in a three-way table with one observation per cell needs to be investigated.

In Chapter 4, joint plots are used as a graphical procedure to diagnose the type of model to fit to three-way arrays of data. The relationship between the various models and the geometrical configurations of the plots allows one to diagnose the type of model which closely fits the data. The application is related to the biplot methods for two-way tables.

One may wonder if the biplot method can be directly used to analyze a three-way table. As the biplot requires data to be organized in matrix form, the three-way tables can only be biplotted if they are collapsed into two-way tables and this may be done by crossing two of the three factors to compromise the rows and isolating the remaining factor levels in the column. Thus, a biplot display of a three-way table will clearly depend upon which two of the factors one chooses to confound.

Besides the biplot, Mandel's interaction plot (Milliken and Johnson, 1989) has been used as a graphical tool for diagnosing a model for two-way tables. This plot would also be

useful and effective especially in showing if some rows or some columns in the data did not fit the model. Some models in the three-factor case may be diagnosed by checking whether each of the interactions is of a particular form using the interaction plot method. Linearity, parallelism, and concurrence in the plot of y_{ijk} against $\hat{\alpha}_i$ for each j and k, y_{ijk} against $\hat{\beta}_j$ for each i and k, and y_{ijk} against $\hat{\gamma}_k$ for each i and j may lead to the diagnosis of regression (linearity on one factor), additivity, and concurrence (of type $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \lambda(\alpha_i\beta_j + \alpha_i\gamma_k + \beta_j\gamma_k) + \lambda^2\alpha_i\beta_j\gamma_k$) models. However, it may be of limited use for smaller data sets with a few special models.

These methods are essentially forcing the three way array into the standard matrix form so that the known methods for two-way data can be applied. But these ignore potentially important three-way interdependencies among the observations since the choice of two-factors from three may affect both fit of the approximation and the form of the interactions. In order to analyze three-way interaction it is recommended to analyze all three factors simultaneously.

The main advantage of the joint plots display of three-way table is that it allows one to investigate all three principal components together at the same time. Furthermore, the methodology for diagnosing the models using the joint plots is rather simple and effective provided there is a fairly good low rank approximation. To interpret the data better, the joint plots together with all other means including the above mentioned graphical techniques and usual analysis of variance should be simultaneously investigated.

The dissertation mostly discussed the geometry of the plots as if the ranks of the related data were exact. Joint plots of data will depart from these ideals but the hope is that the plotted configurations will be sufficiently close to the geometric forms given by the basic models discussed above. The best approximation of Y is always required. If three dimensions seem not to suffice then it may be useful to do one or more additional joint plots on the residuals after eliminating the best dimension-three approximation.

CHAPTER 6. SAS PROGRAM

During this research, several SAS programs are written. Among those, this chapter presents a computer program for the graphical procedure in section 4.4. The program is based on the three-mode principal component analysis. In particular, the program follows the User's guide for TUCKALS3 algorithm by Kroonenberg and Brouwer (1985). Almost the whole program is written by using SAS/IML.

```
OPTIONS LS=75 NOCENTER;

MACRO MODE1 factor1%
MACRO MODE2 factor2%
MACRO MODE3 factor3%
MACRO MODE1_N CYNAME%
MACRO MODE2_N TEMPNAME%
MACRO MODE3_N OXYNAME%


DATA TOXICITY;

INPUT TEMP OXYGEN Y1 Y2 Y3 Y4 Y5;
 CYANIDE=0.16;  Y=Y1;  OUTPUT;
 CYANIDE=0.8;   Y=Y2;  OUTPUT;
 CYANIDE=4.0;   Y=Y3;  OUTPUT;
 CYANIDE=20.0;  Y=Y4;  OUTPUT;
 CYANIDE=100.0; Y=Y5;  OUTPUT;
 DROP Y1-Y5;
```

```
CARDS;
 5 1.5 201 150 131 130 97
 5 3.0 246 164 138 136 102
 5 9.0 271 170 149 127 99
 15 1.5 124 104 86 89 60
 15 3.0 158 111 99 91 74
 15 9.0 207 117 81 87 72
 25 1.5 79 63 50 51 32
 25 3.0 129 54 51 52 46
 25 9.0 142 93 62 51 52
*PROC PRINT;

PROC GLM DATA=TOXICITY;
    CLASS OXYGEN CYANIDE TEMP;
    MODEL Y= TEMP CYANIDE OXYGEN TEMP*CYANIDE TEMP*OXYGEN
             CYANIDE*OXYGEN ;
    OUTPUT OUT=OUTRES P=YHAT R=RESID;
PROC PRINT DATA=OUTRES;
DATA Y_OBS; SET OUTRES;  KEEP Y;

PROC IML;   RESET NOPRINT;
USE Y_OBS;  READ ALL VAR _NUM_ INTO YY;
F1=YY[1:5,]||YY[6:10,]||YY[11:15,];
F2=YY[16:20,]||YY[21:25,]||YY[26:30,];
F3=YY[31:35,]||YY[36:40,]||YY[41:45,];
L=NROW(F1); M=NCOL(F1); N=3; *PRINT L M N;

CYNAME=  {'CY1(.16)' 'CY2(.8)' 'CY3(4)' 'CY4(20)' 'CY5(100)'};
TEMPNAME={'T1(5)' 'T2(15)' 'T3(25)'};
OXYNAME= {'O1(1.5)' 'O2(3)' 'O3(9)'};
STORE L M N CYNAME TEMPNAME OXYNAME;


*================================================;
*           BEGIN MACRO PRINT                    ;
*================================================;
%macro print(FF,row,col);

print &FF[rowname=&row colname=&col format=8.4];
%mend print;
*================================================;
*          END OF MACRO PRINT                    ;
*================================================;


* Fi stands for the ith frontal matrix;
%PRINT(F1,CYNAME,OXYNAME) %PRINT(F2,CYNAME,OXYNAME)
%PRINT(F3,CYNAME,OXYNAME)
```

```
*================================================;
*            BEGIN MACRO ARRAYS                  ;
*================================================;

%MACRO ARRAYS;
L=5; M=3; N=3;

* creating lateral arrays;
do a=1 to m;
   K=F1[,A]||F2[,A]||F3[,A];
  if a=1  then L1=k;
  if a=2  then L2=k;
  if a=3  then L3=k;
end;

* creating horizontal planes ;
do a=1 to l;
   K=L1[A,]//L2[A,]//L3[A,];
  if a=1  then H1=k;
  if a=2  then H2=k;
  if a=3  then H3=k;
  if a=4  then H4=k;
  if a=5  then H5=k;
end;

X=F1||F2||F3;
Y=H1||H2||H3||H4||H5;
Z=L1'||L2'||L3';

%MEND ARRAYS;
*================================================;
*              END MACRO ARRAYS                  ;
*================================================;

%ARRAYS;


sst_x=ssq(x);    * sst(total) using matrix X;
sst=sst_x;       * sst=ss(total)=ssq(x)=ssq(y)=ssq(z);
sst_y=ssq(y);
sst_z=ssq(z);
* print x y z sst_x sst_y sst_z;
STORE X Y Z SST F1 F2 F3 L1 L2 L3 H1 H2 H3 H4 H5;



********* INITIAL COMPONENT MATRICES ************;

%MACRO EIGEN(X,Y,Z);

*obtaining the initial component matrices G0, H0, and E0 from
```

```
 the product matrices P,Q,R;

LOAD MODE1_N MODE2_N MODE3_N CYNAME OXYNAME TEMPNAME;

P_LL=&X*&X';
Q_MM=&Y*&Y';
R_NN=&Z*&Z';

/*computing the eigenvalues and eigenvectors of P,Q,R*/
LP=EIGVAL(P_LL); VP=EIGVEC(P_LL);  LPST=LP/SST;
LQ=EIGVAL(Q_MM); VQ=EIGVEC(Q_MM);  LQST=LQ/SST;
LR=EIGVAL(R_NN); VR=EIGVEC(R_NN);  LRST=LR/SST;

*PRINT"THE EIGENVECTORS AND EIGENVALUES OF THE MODE 1";
COLG={'g1' 'g2' 'g3' 'g4' 'g5'};
COLL={'RAW_LABMDA' 'ST_LAMBDA'};
ROWG=MODE1_N||COLL;
VLP=VP//LP'//LPST';
*%PRINT(VLP,ROWG,COLG);

*PRINT"THE EIGENVECTORS AND EIGENVALUES OF THE MODE 2";
COLH={'h1' 'h2' 'h3'};
COLM={'RAW_MU' 'ST_MU'};
ROWH=MODE2_N||COLM;
VLQ=VQ//LQ'//LQST';
*%PRINT(VLQ,ROWH,COLH);

*PRINT"THE EIGENVECTORS AND EIGENVALUES OF THE MODE 3";
COLE={'e1' 'e2' 'e3' };
COLN={'RAW_NU' 'ST_NU'};
ROWE=MODE3_N||COLN;
VLR=VR//LR'//LRST';
*%PRINT(VLR,ROWE,COLE);

STORE COLG COLH COLE;


BIG_LP=LP[1:2,];   LP_ST=BIG_LP/SST;
BIG_LQ=LQ[1:2,];   LQ_ST=BIG_LQ/SST;
BIG_LR=LR[1:2,];   LR_ST=BIG_LR/SST;

LAMBDA=DIAG(LP_ST');
MU=DIAG(LQ_ST');
NU=DIAG(LR_ST');

TR_LAMB=TRACE(LAMBDA);  TR_LP=SUM(BIG_LP);
TR_MU=TRACE(MU);        TR_LQ=SUM(BIG_LQ);
TR_NU=TRACE(NU);        TR_LR=SUM(BIG_LR);


/* CREATING A SUMMARY TABLE OF EIGENVALUES */
```

```
SUM_RAW=TR_LP//TR_LQ//TR_LR;
SUM_ST=TR_LAMB//TR_MU//TR_NU;
SUM_L=SUM_RAW||SUM_ST;


MATTRIB SUM_L ROWNAME=({'LAMBDA SUM' 'MU SUM' 'NU SUM'})
              label='SUM OF STANDARDIZED COMPONENT WEIGHTS'
              format=8.5;
              PRINT SUM_L;
STORE LAMBDA MU NU;

%MEND EIGEN;
*================================================;
*            END OF MACRO EIGEN                  ;
*================================================;

*PRINT "INITIAL COMPONENT WEIGHTS";
%EIGEN(X,Y,Z);

ESSF_G=SUM(BIG_lp);
ESSF_H=SUM(BIG_lq);
ESSF_E=SUM(BIG_LR);

STORE ESSF_G ESSF_H ESSF_E;



*======== INITIAL COMPONENT MATRICES ============;

S=2; T=2; U=2;
GO=VP[,1:S];
HO=VQ[,1:T];
EO=VR[,1:U];
G=GO; H=HO; E=EO;
STORE G H E;

*================================================;
*          BEGIN MACRO COMPONEN                  ;
*================================================;

%MACRO COMPONEN(X,Y,Z);

/* standardized component matrices */
LOAD L M N LAMBDA MU NU G H E;
g_st=g*LAMBDA##.5;
h_st=h*mu##.5;
E_ST=E*NU##.5;

g_NAME={'g1' 'g2' 'g1_ST' 'g2_ST' };
h_NAMe={'h1' 'h2' 'h1_ST' 'h2_ST' };
```

```
e_NAME={'e1' 'e2' 'e1_ST' 'e2_ST' };


*PRINT"INITIAL COMPONENT LOADINGS OF MODE 1 ";
_GO_=G||G_ST;
*%PRINT(_GO_,MODE1_N,G_NAME);

*PRINT"COMPONENT LOADINGS OF MODE2 " ;
_HO_=H||H_ST;
*%PRINT(_HO_,MODE2_N,H_NAME);

*PRINT"COMPONENT LOADINGS OF MODE3 " ;
_EO_=E||E_ST;
*%PRINT(_EO_,MODE3_N,E_NAME);

%MEND COMPONEN;

*================================================;
*              END OF MACRO COMPONEN              ;
*================================================;


*PRINT "INITIAL COMPONENTS MATRICES";
%COMPONEN(X,Y,Z);

REMOVE LAMBDA MU NU;


************** INITIAL CORE MATRICES **************;

load x y z g h e;
tcf=g`*x*(e@h);
tch=h`*y*(g@e);
tcl=e`*z*(h@g);
s=2; t=2; u=2;

tcoref1=tcf[,1:t];
tcoref2=tcf[,t+1:t+2];
tcoreh1=tch[,1:u];
tcoreh2=tch[,u+1:u+2];
tcorel1=tcl[,1:s];
tcorel2=tcl[,s+1:s+2];

*PRINT " TUCKER'S CORE MATRICES";

gname={'g1' 'g2'};
hname={'h1' 'h2'};
ename={'e1' 'e2'};
```

```
/*
print "frontal plane 1",    tcoref1[rowname=gname colname=hname
                                                  format=8.2];
print "Frontal plane 2",    tcoref2[rowname=gname colname=hname
                                                  format=8.2];
print "Horizontal plane 1", tcoreh1[rowname=hname colname=ename
                                                  format=8.2];
print "Horizontal plane 2", tcoreh2[rowname=hname colname=ename
                                                  format=8.2];
print "Lateral plane 1",    tcorel1[rowname=ename colname=gname
                                                  format=8.2];
print "Lateral plane 2",    tcorel2[rowname=ename colname=gname
                                                  format=8.2];
*/


/* Initial estimated data matrices based on G0,H0,E0 */

x_0=g*tcf*(e`@h`);
y_0=h*tch*(g`@e`);
z_0=e*tcl*(h`@g`);
ssf_x_0=ssq(x_0);



************* MAIN ITERATION OF ALS ****************;

/* PRINT"MAIN ITERATION OF THE TUCKALS3 ALGORITHM"; */
/* Optimization of ssfit */

pa=x*((e*e`)@(h*h`))*x`;
ssfit_0=trace(g`*pa*g);
ssf=ssfit_0; store ssfit_0;
k=2; crit=1; gdiff=1; hdiff=1;  ediff=1;
Do until(crit<.001);
   /* For  gn=pa*g*(g`*pa*pa*g)**(-1/2) */
           gb=(pa*g)`*(pa*g);
           gbp=inv(half(gb));
           gn=pa*g*gbp;  /* note: gbp=(g`*pa*pa*g)**(-1/2) */
           qa=y*((gn*gn`)@(e*e`))*y`;
   /* For  hn=qa*h*(h`*qa*qa*h)**(-1/2) */
           hb=(qa*h)`*(qa*h);
           hbq=inv(half(hb));
           hn=qa*h*hbq;
           ra=z*((hn*hn`)@(gn*gn`))*z`;
   /* For  en=ra*e*(e`*ra*ra*e)**(-1/2) */
           eb=(ra*e)`*(ra*e);
           ebr=inv(half(eb));
           en=ra*e*ebr;
           pn=x*((en*en`)@(hn*hn`))*x`;
```

```
            ssfn=trace(gn`*pn*gn);

            crit=abs(ssfn-ssf);
               gdif=sqrt( sum( (gn-g)#(gn-g) ) );
               hdif=sqrt( sum( (hn-h)#(hn-h) ) );
               edif=sqrt( sum( (en-e)#(en-e) ) );

               kk=kk//k;
               ssfit=ssfit//ssfn;
               increm=increm//crit;
               gdi=gdi//gdif;
               hdi=hdi//hdif;
               edi=edi//edif;

         g=gn; h=hn; e=en;
         pa=pn; ssf=ssfn;
         k=k+1;
end;


/* final core matrices */

cf=g`*x*(e@h);
ch=h`*y*(g@e);
cl=e`*z*(h@g);

STORE K G H E CL CF CH;



******** UNSTANDARDIZED CORE MATRICES *************;

s=2; t=2; u=2;
coref1=cf[,1:t];
coref2=cf[,t+1:t+2];
coreh1=ch[,1:u];
coreh2=ch[,u+1:u+2];
corel1=cl[,1:s];
corel2=cl[,s+1:s+2];

store coref1 coref2 coreh1 coreh2 corel1 corel2;


********** HISTORY OF ITERATION STEPS ************;

/*
PRINT K KK SSFIT INCREM GDI HDI EDI;
PRINT"STEPS OF MAIN ITERATION OF ALS ";
k_=1; kk=k_//kk; ssfit= ssfit_0//ssfit; dummy_=1; increm=dummy_//increm;
gdi=dummy_//gdi; hdi=dummy_//hdi; edi=dummy_//edi;
```

```
SUMM=KK||SSFIT||INCREM||GDI||HDI||EDI ;
CREATE SUMMARY FROM SUMM;
APPEND FROM SUMM;

DATA SUMMARY;
     set summary;
     rename col1=k col2=ssfit col3=crit
            col4=gdiff col5=hdiff col6=ediff;
proc print data=summary;
*/



**************** COMPONENT WEIGHTS ******************;

xhat=g*cf*(e'@h');
yhat=h*ch*(g'@e');
zhat=e*cl*(h'@g');
store xhat yhat zhat;


*PRINT "EIGENVALUES AFTER THE ITERATION";

%eigen(xhat,yhat,zhat);



*************** FINAL COMPONENT MATRICES ************;

load l m n;
g_st=g*lambda##.5;              /* standardized component weights */
h_st=h*mu##.5;
e_st=e*nu##.5;
store h_st ;
eh=(e*e')@(h*h');


*PRINT"FINAL COMPONENTS MATRICES";

*PRINT"COMPONENT LOADINGS OF MODE 1";
G_NAME={'g1' 'g2' 'g1_ST' 'g2_ST' };
H_NAME={'h1' 'h2' 'h1_ST' 'h2_ST' };
E_NAME={'e1' 'e2' 'e1_ST' 'e2_ST' };

LOAD MODE1_N MODE2_N MODE3_N;


*PRINT"COMPONENT LOADINGS OF MODE1 ";
_G_=G||G_ST;
*%PRINT(_G_,MODE1_N,G_NAME);
```

```
*PRINT"COMPONENT LOADINGS OF MODE2 ";
_H_=H||H_ST;
*%PRINT(_H_,MODE2_N,H_NAME);

*PRINT"COMPONENT LOADINGS OF MODE3 ";
_E_=E||E_ST;
*%PRINT(_E_,MODE3_N,E_NAME);



************ SUMMARY OF ITERATION ***************;

/* computing the fitted sum of squares   */

load xhat cf cl ch;
ssf_g=ssq(xhat);        /* ss(fit) based on the g,h,e */
ssq_cf=ssq(cf);         /* ssq(frontal core plane)=ss(fit) */
ssq_t=trace(cl*cl');    /* trace(cc')=ss(fit) */
ssfit=ssf_g;
ss_res=sst-ssfit;

/* differences between ss(fit) and the estimated ssf_g  */

difit_g=essf_g-ssfit;
difit_h=essf_h-ssfit;
difit_e=essf_e-ssfit;
Incr_ssf=ssfit-ssfit_0;  *ssf(converged)-ssf(initial estimate);

/* summmary of iteration */

K_=K-1; SSFIT_=SSFIT; KKS=0;
RESULT1=K_//SST//SSFIT_0//SSFIT_//SS_RES//essf_g//essf_h//essf_e//
        DIFIT_G//DIFIT_H//DIFIT_E//INCR_SSF;
RESULT2=KKS//SST//SSFIT_0//SSFIT_//SS_RES//essf_g//essf_h//essf_e//
        DIFIT_G//DIFIT_H//DIFIT_E//INCR_SSF;
RESULT3=RESULT2/SST;
RESULT=RESULT1||RESULT3;

mattrib result rowname=({ 'number of iteration'
                          'SS(TOTAL)'
                          'SS(FIT) before iteration'
                          'SS(FIT) after iteration' 'SS(RESIDUAL)'
                          'estimated SS(fit) based on G'
                          'estimated SS(fit) based on H'
                          'estimated SS(fit) based on E'
                          'est.SS(fit) on G - SS(fit)'
                          'est.SS(fit) on H - SS(fit)'
                          'est.SS(fit) on E - SS(fit)'
                          'Increment in fit' })
               label='SUMMARY OF ITERATION  ' format=8.5;
```

```
                print result;

 STORE SSFIT;    /* PRINT SSFIT; */


************** SCALED CORE MATRICES **************;


LOAD COREF1 COREF2 COREH1 COREH2 COREL1 COREL2  LAMBDA MU NU L M N;

/*compute ss(total) scaled matrices */

cf_st=cf#cf/sst;
cf_st1=cf_st[,1:2];
cf_st2=cf_st[,3:4];
ch_st=ch#ch/sst;
ch_st1=ch_st[,1:2];
ch_st2=ch_st[,3:4];
cl_st=cl#cl/sst;
cl_st1=cl_st[,1:2];
cl_st2=cl_st[,3:4];

/* PRINT"CORE MATRICES";

gname={'g1' 'g2'};
hname={'h1' 'h2'};
ename={'e1' 'e2'};

print "Frontal plane 1", coref1[rowname=gname colname=hname format=8.2];
print "Frontal plane 2", coref2[rowname=gname colname=hname format=8.2];
print"Horizontal plane 1",coreh1[rowname=hname colname=ename format=8.2];
print"Horizontal plane 2",coreh2[rowname=hname colname=ename format=8.2];
print "Lateral plane 1", corel1[rowname=ename colname=gname format=8.2];
print "Lateral plane 2", corel2[rowname=ename colname=gname format=8.2];
*/

CF_1=COREF1||CF_ST1;  CF_2=COREF2||CF_ST2;
CH_1=COREH1||CH_ST1;  CH_2=COREH2||CH_ST2;
CL_1=COREL1||CL_ST1;  CL_2=COREL2||CL_ST2;
gname={'g1' 'g2' 'g1_st' 'g2_st' };
hname={'h1' 'h2' 'h1_st' 'h2_st' };
ename={'e1' 'e2' 'e1_st' 'e2_st' };

/*
PRINT "FRONTAL PLANE 1",    CF_1[ROWNAME=GNAME COLNAME=HNAME FORMAT=8.2];
PRINT "FRONTAL PALNE 2",    CF_2[ROWNAME=GNAME COLNAME=HNAME FORMAT=8.2];
PRINT "HORIZONTAL PLANE 1", CH_1[ROWNAME=HNAME COLNAME=ENAME FORMAT=8.2];
PRINT "HORIZONTAL PLANE 2", CH_2[ROWNAME=HNAME COLNAME=ENAME format=8.2];
PRINT "LATERAL PLANE 1",    CL_1[ROWNAME=ENAME COLNAME=GNAME format=8.2];
PRINT "LATERAL PLANE 2",    CL_2[ROWNAME=ENAME COLNAME=GNAME format=8.2];
*/
```

```
** GH-JOINT PLOT **;


******* INNER PRODUCTS FOR 1ST GH-JOINT PLOT *******;

proc iml; reset noprint;
rowmark={'g1' 'g2' 'g3' 'g4' 'g5' };
colmark={'h1' 'h2' 'h3'};
LOAD L M N COREF1 COREF2 G H E cyname oxyname tempname;
CALL SVD(u,q,v,coref1);
CONST=(L/M)##(1/4);
d=diag(q);
gstar=const*g*u*sqrt(d);
hstar=(1/const)*h*v*sqrt(d);

*inner products;
y_f1=gstar*hstar';
INNER_1=G*COREF1*H';
*PRINT Y_F1[ROWNAME=CYNAME COLNAME=OXYNAME FORMAT=8.4];


*************** 1ST GH_JOINT PLOT ****************;

GHSTAR=GSTAR//HSTAR; *PRINT GHSTAR;
LABEL_V=CONCAT('*',COLMARK);
LABEL=rowmark||LABEL_V;
*PRINT "FIRST M SCORES OF MODE2 FOLLOWING L SCORES OF MODE 1";
PRINT GHSTAR[ROWNAME=LABEL FORMAT=8.4];

XY=GHSTAR[1:L+M,1:2];
ID=rowmark'//LABEL_V';
XLABEL='COMPONENT 1';
YLABEL='COMPONENT 2';
TITLE='1ST GH-JOINT PLOT'; *OK?;
OPTIONS LS=75;
call pgraf(xy,id,xlabel,ylabel,title);



******* INNER PRODUCTS FOR 2ND GH-JOINT PLOT ********;


LOAD L M N COREF1 COREF2 G H E;
CALL SVD(U,Q,V,COREF2);
CONST=(L/M)##(1/4);
d=diag(q);
wt_g=const*sqrt(d);
gstar=const*g*u*sqrt(d);
```

```
hstar=(1/const)*h*v*sqrt(d);


*inner products = Gstar*Hstar' = G*COREF1*H'= Y_F1 ;
y_f2=gstar*hstar';
INNER_2=G*COREF2*H';
*PRINT Y_F2[ROWNAME=cyname COLNAME=oxyname FORMAT=8.4];



*************** 2ND GH_JOINT PLOT ****************;


GHSTAR=GSTAR//HSTAR;  *PRINT GHSTAR;
LABEL_V=CONCAT('*',colmark);
LABEL=rowmark||label_v;
*PRINT "FIRST M SCORES OF MODE2 FOLLOWING L SCORES OF MODE 1";
PRINT GHSTAR[ROWNAME=LABEL FORMAT=8.4];

xy=ghstar[1:l+m,1:2];
ID=rowmark'//label_v';
XLABEL='COMPONENT 1';
YLABEL='COMPONENT 2';
TITLE='2nd GH-joint plot ';
OPTIONS LS=75;
call pgraf(xy,id,xlabel,ylabel,title);
```

BIBLIOGRAPHY

Bradu, D. and Gabriel, K. R. (1978). The biplot as a diagnostic tool for models of two-way tables. *Technometrics,* **20,** 47-68.

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.

Corsten, L. C. A. and Van Eijnsbergen, A. C. (1972). Multiplicative effects in two-way and analysis of variance. *Statistica Neerlandica,* **26,** 61-68.

Edgington E. S. (1980). *Randomization Tests.* Marcel Dekker, New York, 1-287.

Fisher, R. A. and MacKenzie, W. A. (1923). Studies in crop variation II. The manurial response of different potato varieties. *Journal of Agricultural Science,* **13,** 311-20

Gabriel, K. R. (1971). The biplot graphic display of matrices with applications to principal components. *Biometrika,* **58,** 453-67.

Gabriel, K. R. (1978). Least square approximation of matrices by additive and multiplicative models. *Journal of the Royal Statistical Society, series B,* **40,** 186-196.

Gabriel K. R. (1980). *Interpreting Multivariate Data,* editor V. Barnett. Wiley, 147-173.

Gollob, H. F. (1968). A statistical model which combines features of factor analytic and analysis of variance techniques. *Psychometrika,* **33,** 73-116.

Gower, J. C. (1990). Three-dimensional biplots. *Biometrika,* **77,** 773-785

Gower, J. C. and Digby, P. G. N. (1981). Expressing complex relationships in two dimensions. In *Interpreting Multivariate Data*, editor V. Barnett. Wiley, New York, 83-118.

Gower, J. C. and Harding, S. A. (1988). Nonlinear biplots. *Biometrika* **75**, 445-55.

Harshman, R. A. (1970). Foundations of the PARAFAC procedure: Models and conditions for an exploratory multi-modal factor analysis. *UCLA Working papers in Phonetics,* **16,** 1-84.

Hartley, H. O. (1977). Statistical methods for Digital Computers. Wiley, New York, 16-34

Jackson, J. E. (1991). *A User's Guide to Principal Components.* Wiley, New York.

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

Jolliffe, I. T. (1986). *Principal Component Analysis.* Springer-Verlag, New York.

Kester, N. K. (1979). Diagnosing and fitting concurrent and related models for two-way and higher way layouts. Unpublished Ph.D. thesis, University of Rochester.

Kettenring, J. R. (1983a). Components of interaction in analysis of variance models with no replications. In *Contributions to Statistics: Essays in Honor of Norman L. Johnson*, editor P. K. Sen. North-Holland, Amsterdam, 283-297.

Kettenring, J. R. (1983b). A case study in data analysis. *Proceedings of Symposia in Applied Mathematics*, **28,** 105-139.

Krishnaiah P. R. and Yochmowitz, M. G. (1980). Inference on the structure of interaction in two-way classification model. In *Handbook of Statistics*, editor P. R. Krishnaiah, North-Holland Publishing Company, 973-994.

Kroonenberg, P. M. (1983). *Three-Mode Principal Component Analysis.* DSWO press, Leiden.

Kroonenberg, P. M. (1989). Singular value decomposition of interaction in three-way contingency tables. In *Multiway Data Analysis*, editors R. Coppi and S. Bolasco. Elsevier, Amsterdam, 169-184.

Kroonenberg, P. M. and Brouwer P. (1985). *User's Guide to TUCKALS3, Version 4.0.* WR Section W.E.P. University of Lieden, Netherlands.

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

Kroonenberg, P. M. and Van der Voort, T. H. A. (1987). Multiplicative decompositie van interacties bij oordelen over de werkelijkheidswaarde van televisiefilms [Multiplicative decomposition of interactions for judgements of realism in television films]. *Kwantitatieve Methoden,* **8 (23)**, 117-144.

Kruskal, J. B. (1989). Rank, decomposition, and uniqueness for 3-way and n-way arrays. In *Multiway Data Analysis*, editors R. Coppi and S. Bolasco. Elsevier, Amsterdam, 7-18.

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

Mandel, J. (1969). The partitioning of interaction in analysis of variance. *Journal of Research of the National Bureau of Standards*, **73B**, 309-328.

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

Manly B. F. J. (1991). *Randomization and Monte Carlo Methods in Biology.* Chapman and Hall, New York, 1-281.

Marasinghe, M. G. and Johnson, D. E. (1981). Testing subhypotheses in the multiplicative interaction model. *Technometrics*, **23**, 385-393.

Milliken, G. A. and Johnson D. E. (1989). *Analysis of Messy Data, Vol 2: Nonreplicated Experiments.* Van Nostrand Reinhold, New York, 1-199.

Penrose R. (1956). On the best approximate solutions of linear matrix equations. *Proceedings of the Cambridge Philosophical Society,* **51**, 406-413.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, second edition. Wiley, New York.

Rutishauser, H. (1969). Computational aspects of F. L. Bauer's simultaneous iteration method. *Numerische Mathematik,* **13**, 4-13.

Seber, G. A. F. (1984). *Multivariate Observations.* Wiley, New York. 1-686.

Snee R. D. (1982) . Nonadditivity in a two-way classification: Is it interaction or nonhomogeneous variance? *Journal of the American Statistical Association,* **77,** 515-519.

Schuurmann, F. J., Krishnaiah, P. R. and Chattopadhyah, 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.

Tucker, L. R. (1963). Implications of factor analysis of three-way matrices for measurement of change. In *Problems in Measuring Change*, editor C. W. Harris. University of Wisconsin Press, Madison, Wisconsin, 122-137.

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.

Wilks, S. S. (1962). *Mathematical Statistics.* Wiley, New York, 1-644 (568).

Wuhrmann, K. and Worker, H. (1953). Über die Giftwirkungen von Ammoniak und Zyanidlösungen mit verschiedner Sauerstoffspannung und Temperatur auf Fische. Schwiez. Z. *Hydrol.* **15**, 235-260.

Yochmowitz, M. G. and Cornell, R. G. (1978). Stepwise tests for multiplicative components of interaction. *Technometrics*, **20**, 79-84

Young F. W., De Leeuw, J., and Takane, Y. (1980). Quantifying qualitative data. In *Similarity and Choice* (E. D. Lantermann and H. Feger eds.). Hans Huber: Wein, Austria

APPENDIX A. Proof of $||D||^2 = ||C||^2$

Suppose $s = l$, $t = m$, and $u = n$. Then

$$||D||^2 = \sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}^2 = \sum_{p=1}^{s}\sum_{q=1}^{t}\sum_{r=1}^{u} \hat{c}_{pqr}^2 = ||C||^2.$$

Proof: From the equations (3.7) and (3.9), we have

$$(1)\quad d_{ijk} = \sum_{p=1}^{l}\sum_{q=1}^{m}\sum_{r=1}^{n} \hat{c}_{pqr}\, \hat{g}_{ip}\, \hat{h}_{jq}\, \hat{e}_{kr} \tag{A.1}$$

$$(2)\quad \sum_{i=1}^{l} \hat{g}_{ip} = \sum_{j=1}^{m} \hat{h}_{jq} = \sum_{k=1}^{n} \hat{e}_{kr} = 0\ ,\quad \forall\ p,\ q,\ r. \tag{A.2}$$

$$\sum_{i=1}^{l} \hat{g}_{ip}^2 = \sum_{j=1}^{m} \hat{h}_{jq}^2 = \sum_{k=1}^{n} \hat{e}_{kr}^2 = 1\ ,\quad \forall\ p,\ q,\ r. \tag{A.3}$$

$$\sum_{i=1}^{l} \hat{g}_{ip}\hat{g}_{ip'} = \sum_{j=1}^{m} \hat{h}_{jq}\hat{h}_{jq'} = \sum_{k=1}^{n} \hat{e}_{kr}\hat{e}_{kr'} = 0\ ,\quad \forall\ p \neq p'\ ,\ q \neq q',\ r \neq r'. \tag{A.4}$$

By using these identities and switching the order of summations if it is necessary, we get

$$
\begin{aligned}
||D||^2 &= \sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}^2 \\
&= \sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} \Big(\sum_{p=1}^{l}\sum_{q=1}^{m}\sum_{r=1}^{n} \hat{c}_{pqr}\, \hat{g}_{ip}\, \hat{h}_{jq}\, \hat{e}_{kr} \Big)^2 \qquad \text{by (A.1)} \\
&= \sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} \Big\{ \sum_{p=1}^{l}\sum_{q=1}^{m}\sum_{r=1}^{n} \Big(\hat{c}_{pqr}\, \hat{g}_{ip}\, \hat{h}_{jq}\, \hat{e}_{kr} \Big)^2 \Big\} \quad \text{by (A.2) and (A.4)} \\
&= \sum_{i=1}^{l}\sum_{j=1}^{m} \Big(\sum_{p=1}^{l}\sum_{q=1}^{m}\sum_{r=1}^{n} \hat{c}_{pqr}\, \hat{g}_{ip}\, \hat{h}_{jq} \Big)^2 \qquad \text{by (A.3)} \\
&= \sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} \hat{c}_{pqr}^2 \qquad \text{by (A.3)} \\
&= ||C||^2
\end{aligned}
$$

□

APPENDIX B. Likelihood ratio statistic for group testing

In Section 3.4, the likelihood ratio test statistic Λ is derived for testing the hypothesis H_0: $c^2 = 0$ versus H_1: $c^2 > 0$. In this appendix, we consider general hypothesis H_0: $c^2_{111} = \cdots = c^2_{stu} = 0$ versus H_1: at least one of $c^2_{pqr} \neq 0$, for $1 \leq p \leq s$, $1 \leq q \leq t$, and $1 \leq r \leq u$.

Suppose Y is an $l \times m \times n$ array of mutually independent normal random variables all with the same variance σ^2 such that μ^{ω}_{ijk}, the expected value of y_{ijk} under H_0, is

$$E(y_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk},$$

or equivalently,

$$E(y_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_k + \tau_{ij} + \rho_{ik} + \phi_{jk}.$$

with the standard meaning and constraints on parameters. A test is required ascertaining possible three-factor interaction terms of multiplicative type, i.e., μ^{Ω}_{ijk}, the expected value of y_{ijk} under $H_0 \cup H_1$, is

$$E(y_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_k + \tau_{ij} + \rho_{ik} + \phi_{jk} + \eta_{ijk}.$$

where $\eta_{ijk} = \sum_{i=1}^{l} \sum_{j=1}^{m} \sum_{k=1}^{n} c_{pqr}\, g_{ip}\, h_{jq}\, e_{kr}$. In addition to the standard constraints on parameters, the condition $\sum_i \eta_{ijk} = \sum_j \eta_{ijk} = \sum_k \eta_{ijk} = 0$ are imposed.

A likelihood ratio is based on the lower critical region for the following statistic to the power $\frac{lmn}{2}$:

$$\Lambda^* = \frac{L(\hat{\omega})}{L(\hat{\Omega})} = \frac{\min_{\underline{\theta}\,\omega} \; || \, Y\text{-}E(Y_\Omega) \, ||^2}{\min_{\underline{\theta}\,\Omega} \; || \, Y\text{-}E(Y_\omega) \, ||^2}$$

$$= \frac{\min_{\underline{\theta}\,\omega} \; \sum_{i=1}^{l} \sum_{j=1}^{m} \sum_{k=1}^{n} (y_{ijk} - \mu_{ijk}^{\Omega})^2}{\min_{\underline{\theta}\,\Omega} \; \sum_{i=1}^{l} \sum_{j=1}^{m} \sum_{k=1}^{n} (y_{ijk} - \mu_{ijk}^{\omega})^2}$$

$$= \frac{\min_{\{c,\, g,\, h, e\}} \; \sum_{i=1}^{l} \sum_{j=1}^{m} \sum_{k=1}^{n} (d_{ijk} - \eta_{ijk})^2}{\sum_{i=1}^{l} \sum_{j=1}^{m} \sum_{k=1}^{n} d_{ijk}^{\,2}}$$

$$= \frac{\sum_{i=1}^{l} \sum_{j=1}^{m} \sum_{k=1}^{n} (d_{ijk} - \hat{\eta}_{ijk})^2}{\sum_{i=1}^{l} \sum_{j=1}^{m} \sum_{k=1}^{n} d_{ijk}^{\,2}} \qquad \text{(B.1)}$$

where the parameters under $H_0 \cup H_1$ are $\underline{\theta}_\Omega = [\, \mu, \; \underline{\alpha}, \; \underline{\beta}, \; \underline{\gamma}, \; \underline{\tau}, \; \underline{\rho}, \; \underline{\phi}, \; \underline{c}, \; \underline{g}, \; \underline{h}, \; \underline{e}, \; \sigma^2 \,]$, and under H_0 are $\underline{\theta}_\omega = [\, \mu, \; \underline{\alpha}, \; \underline{\beta}, \; \underline{\gamma}, \; \underline{\tau}, \; \underline{\rho}, \; \underline{\phi}, \; \sigma^2 \,]$ as defined in Section 3.4. Also d_{ijk}, defined as in (3.6), is the residual with respect to the maximum likelihood estimates of μ_{ijk} under H_0 and

$$\eta_{ijk} = \sum_{p=1}^{l} \sum_{q=1}^{m} \sum_{r=1}^{n} c_{pqr} \, g_{ip} \, h_{jq} \, e_{kr},$$

$$\hat{\eta}_{ijk} = \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} \hat{c}_{pqr}\, \hat{g}_{ip}\, \hat{h}_{jq}\, \hat{e}_{kr}.$$

The numerator of Λ^* becomes

$$\sum_{i=1}^{l} \sum_{j=1}^{m} \sum_{k=1}^{n} (d_{ijk} - \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} \hat{c}_{pqr}\, \hat{g}_{ip}\, \hat{h}_{jq}\, \hat{e}_{kr})^2$$

$$= \sum_{i=1}^{l} \sum_{j=1}^{m} \sum_{k=1}^{n} d_{ijk}^2 + \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} (\hat{c}_{pqr}\, \hat{g}_{ip}\, \hat{h}_{jq}\, \hat{e}_{kr})^2$$

$$- 2 \sum_{i=1}^{l} \sum_{j=1}^{m} \sum_{k=1}^{n} d_{ijk} \left(\sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} \hat{c}_{pqr}\, \hat{g}_{ip}\, \hat{h}_{jq}\, \hat{e}_{kr} \right)$$

$$= \sum_{i=1}^{l} \sum_{j=1}^{m} \sum_{k=1}^{n} d_{ijk}^2 + \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} \hat{c}_{pqr}{}^2 - 2 \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} \hat{c}_{pqr}^2$$

$$= \sum_{i=1}^{l} \sum_{j=1}^{m} \sum_{k=1}^{n} d_{ijk}^2 - \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} \hat{c}_{pqr}^2$$

From the second line to third line, we note that the cross product term is as follow:

$$\sum_{i=1}^{l} \sum_{j=1}^{m} \sum_{k=1}^{n} d_{ijk} \left(\sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} \hat{c}_{pqr}\, \hat{g}_{ip}\, \hat{h}_{jq}\, \hat{e}_{kr} \right)$$

$$= \sum_{i=1}^{l} \sum_{j=1}^{m} \sum_{k=1}^{n} \left[\left(\sum_{p=1}^{l} \sum_{q=1}^{m} \sum_{r=1}^{n} \hat{c}_{pqr}\, \hat{g}_{ip}\, \hat{h}_{jq}\, \hat{e}_{kr} \right) \left(\sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} \hat{c}_{pqr}\, \hat{g}_{ip}\, \hat{h}_{jq}\, \hat{e}_{kr} \right) \right]$$

$$= \sum_{i=1}^{l} \sum_{j=1}^{m} \sum_{k=1}^{n} \left(\sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} \hat{c}_{pqr}\, \hat{g}_{ip}\, \hat{h}_{jq}\, \hat{e}_{kr} \right)^2 + \sum_{p=s+1}^{l} \sum_{q=t+1}^{m} \sum_{r=u+1}^{n} \hat{c}_{pqr}\, \hat{g}_{ip}\, \hat{h}_{jq}\, \hat{e}_{kr})$$

$$= \sum_{p=1}^{s}\sum_{q=1}^{t}\sum_{r=1}^{u} \hat{c}_{pqr}^2 \qquad \text{since the second term is zero (cf. Appendix A).}$$

As a result, the likelihood ratio test statistic becomes

$$\Lambda^{**} = (\Lambda^{*})^{\frac{2}{lmn}} = \frac{\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n}\left(d_{ijk} - \sum_{p=1}^{s}\sum_{q=1}^{t}\sum_{r=1}^{u} \hat{c}_{pqr}\, \hat{g}_{ip}\, \hat{h}_{jq}\, \hat{e}_{kr}\right)^2}{\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}^{\,2}}$$

$$= \frac{\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}^2 - \sum_{p=1}^{s}\sum_{q=1}^{t}\sum_{r=1}^{u} \hat{c}_{pqr}^2}{\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}^2}$$

In this case, the hypothesis H_0: $c_{111}^2 = \cdots = c_{stu}^2 = 0$ is rejected according as

$$\Lambda^{**} < k^*$$

or equivalently,

$$\Lambda = 1 - \Lambda^{**} = \frac{\sum_{p=1}^{s}\sum_{q=1}^{t}\sum_{r=1}^{u} \hat{c}_{pqr}^2}{\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}^2} > k \qquad \text{(B.2)}$$

where k is such that

$$P(\Lambda > k \mid H_0) = \alpha.$$

When the test concerns only one multiplicative term, then the test statistic Λ becomes

$$\Lambda_1 = \frac{\hat{c}^2}{\sum_{i=1}^{l}\sum_{j=1}^{m}\sum_{k=1}^{n} d_{ijk}^{\,2}} \tag{B.3}$$

where $\hat{c} = \max\{\, \hat{c}^2_{pqr} \mid 1 \le p \le s,\ 1 \le q \le r,\ 1 \le r \le u\}$ as defined in Section 3.4. Notice that this statistic Λ_1 is the same as the statistic Λ (3.38) in Section 3.4.2.

VITA

of

Kyoungah S. See

Education

Ph. D.- May 1993	Virginia Polytechnic Institute and State University Blacksburg, Virginia Dissertation: Three-mode principal component analysis in designed experiments
M. S. - June 1989	The Ohio State University Columbus, OH
B. S. - June 1985	The Ohio State University Columbus, OH

Teaching Experience

Graduate Teaching Associate:	Department of Statistics Virginia Polytechnic Institute and State University August 1989 - May 1989
Graduate Teaching Assistant:	Department of Statistics The Ohio State University August 1985 - March 1989
Senior Instructional Aide:	Department of Mathematics The Ohio State University August 1984 - June 1985

Scholarly Activity

Technical report:	Kyoungah See and Eric P. Smith "Diagnosing models for three-way data arrays using joint plots" *Technical report* number **92-2**, Department of Statistics, Virginia Tech, May 1992
Paper submitted:	K. See and E. P. Smith "Model Assessment for three-way data arrays using joint plots" submitted, July 1992
Contributed Talk:	"A new graphical procedure for model building in a three-way table", The Biometrics Joint Meeting, Philadelphia, Pennsylvania, March 22, 1993
	"A new test for three-way interactions in ANOVA", American Statistical Association Winter Conference, Ft. Lauderdale, Florida, January 4, 1993
	"Diagnosing models for three-way data analysis using joint plots", American Statistical Association Annual Conference, Boston Massachusetts, August 1992

Awards

Student Travel Fellowship Award:

ENAR (Eastern North American Region) Student Travel Fellowship for Joint Meeting of the Biometric Society, Philadelphia, Pennsylvania, March 1993

Graduate Teaching Excellence Awards:

1987 - 1988	Teaching Excellence Award for the academic year 1986 – 87 The Ohio State University
1986 - 1987	Teaching Excellence Award for the academic year 1987 – 88. The Ohio State University