

The TUCKALS line

A suite of programs for three-way data analysis

Pieter M. Kroonenberg

Department of Education, Leiden University, Leiden, The Netherlands

Abstract: This paper describes two programs (TUCKALS2 en TUCKALS3) with which three-way data can be analysed. Both are based on generalisations of standard (two-way) principal component analysis. The working of the programs, and the basic theory behind them, is explained, and is illustrated with data on the influence of alcohol on the behaviour of Australian twins.

Keywords: Three-mode principal component analysis; Three-way data; Twins; Computer programs.

1. Introduction

Three-mode principal component analysis using a model, which will be referred to as the Tucker3 model, was first formulated within the context of the social sciences by Tucker (1963), and in subsequent papers Tucker (1964, 1966) refined especially its mathematical description. In the latter paper, Tucker also proposed several methods to solve the estimation of the parameters. A stochastic version of this model was first proposed by Bloxom (1968), and was further developed by Bentler and Lee (1978, 1979), Lee and Fong (1983). Kroonenberg and De Leeuw (1980) presented an improved, least squares solution for the original Tucker3 model. In Kroonenberg (1983a) an overview is presented of the state of the art up to that moment, and several later developments are contained in the readers edited by Law et al. (1984), and Coppi and Bolasco (1989), and review papers by Geladi (1989), Kroonenberg (1992), and Smilde (1992).

Correspondence to: P.M. Kroonenberg, Dept. of Education, Leiden University, Wassenaarseweg 52, 2333 AK Leiden, The Netherlands.

Also treated in this paper, is the Tucker2 model, which is less restricted than the Tucker3 model. Tucker (1975) was probably the first to formulate it explicitly, but a slightly less general version was proposed by other authors (Israelsson, 1969; Carroll and Chang, 1972; Jennrich, 1973). The Tucker2 model belongs to the class of individual differences models of which it is the most general representative. It has also been called a 'generalized subjective metrics model' (Sands and Young, 1980, p. 41). The other similar models have mainly been developed within the context of multidimensional scaling. General discussions of individual differences models and their interrelationships can, for instance, be found in Arabie et al. (1987) and their references. The Tucker2 model and its estimation have been fully described by Kroonenberg and De Leeuw (1977), and partially in Kroonenberg and De Leeuw (1980), and Kroonenberg (1983a). Hierarchies of three-way models from both the French and Anglo-Saxon literature which include the Tucker3 and Tucker2 models respectively, have been presented by Kiers (1988, 1991).

2. Three-way data

Consider the situation in which a number of persons have rated twenty abstract paintings using some ten different rating scales, which measure the feelings these paintings elicit. Suppose a researcher wants to know (1) if there is a common structure underlying the usage of the rating scales with respect to the paintings, (2) how the various subjects perceive this common structure, and/or (3) whether subjects can be seen as types or combinations of types in their use of the rating scales for the pictures. Although all subjects might agree on the dimensions of feelings elicited by the paintings, for some subjects certain dimensions might be more important and/or more correlated than for other subjects, and one could imagine that different types of subjects evaluate paintings in a different way. A way to gain insight in such problems is to determine the (low)dimensional structures for paintings, rating scales, and subjects expressed in components, and combining these component spaces in some way to assess the relationships between components. The data for this example can be arranged into a three-dimensional block of variables by conditions by subjects. Such a block is generally referred to as a *three-way data matrix*. We will use the word *way* to indicate a collection of indices by which the data can be classified, while the word *mode* will indicate the entities that make up one of ways of the data box, here paintings, rating scales and subjects. Thus data boxes always have three ways, but there are only three modes when each of the three ways consist of different entities. Sets of correlation matrices have therefore two modes, while our example has three modes (see Carroll and Arabie, 1980, for further data types).

The first way (variables) has index i running along the vertical axis, the second way (conditions) has index j running along the horizontal axis, and the third way (subjects) has index k running along the 'depth' axis of the box. The

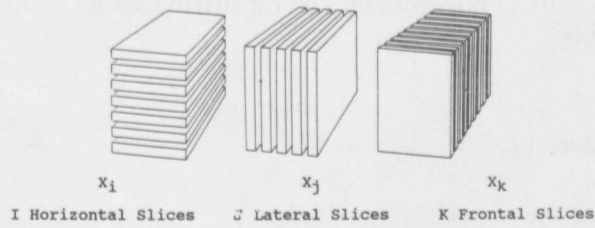


Fig. 1. Slices, the two-way submatrices of the three-way matrix X .

number of levels in each way is I , J , and K respectively, while the number of components will be indicated with P , Q , and R , respectively, with p , q , and r the corresponding indices. The component matrices will be A , B , and C , and the core matrix G . The $I \times J \times K$ three-way data matrix X is thus defined as the collection of elements:

$$\{x_{ijk} \mid i = 1, \dots, I; j = 1, \dots, J; k = 1, \dots, K\}.$$

A three-way matrix can also be seen as a collection 'normal' (= two-way) matrices or slices. There are three different arrangements for this, as is shown in Figure 1. Furthermore, one can break up a three-way matrix into one-way submatrices (or vectors), called fibers (see Figure 2). The slices will be called frontal slices, horizontal slices, and lateral slices. The fibers will be called rows, columns, and tubes. The terminology used here is largely based on Harshman and Lundy (1984a).

Note that in most multivariate statistical models, subjects are considered a random factor, but in the data-analytic models such as those considered in this paper this is not necessarily the case. Notwithstanding, in many applications the status of the subject mode is somewhat different, because after all they are the 'data generators'. In addition, three-way models are sometimes used to establish whether it is reasonable to treat the subjects as replications, rather than members of different subsamples. Other applications are true population studies and no true stochastic framework can be formulated, for instance in the analysis

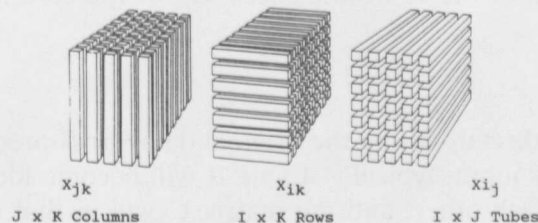


Fig. 2. Fibers, the one-way submatrices of the three-way matrix X .

of the performance of genotypes on several attributes at various locations (e.g. Basford et al., 1990).

3. Model description

Tucker3 model

The Tucker3 model is the factorization of the three-way data matrix $X = \{x_{ijk}\}$, such that

$$x_{ijk} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R a_{ip} b_{jq} c_{kr} g_{pqr} + e_{ijk}$$

$$(i = 1, \dots, I; j = 1, \dots, J; k = 1, \dots, K) \quad (1)$$

where the coefficients a_{ip} , b_{jq} and c_{kr} are the elements of the component matrices A , B , and C respectively, the g_{pqr} are the elements of the three-way core matrix G , and the e_{ijk} are the errors of approximation collected in the three-way matrix E . A is the $(I \times P)$ matrix with the coefficients of the variables of the first mode on the variable components. B is the $(J \times Q)$ matrix with coefficients of the conditions, and C the $(K \times R)$ coefficient matrix of the subjects. In the original data matrix X every element of the matrix represents the value of a specific combination of levels of the original modes. In a similar manner each element of the core matrix represents the value or weight of a specific combination of the components of the modes.

A matrix formulation of the model is

$$X = AG(C' \otimes B') + E \quad (2)$$

where X , E , and G are written as ordinary two-way matrices of order $(I \times JK)$, $(I \times JK)$ and $(P \times QR)$ by making use of so-called combination modes (Tucker, 1966, p. 281), and \otimes denotes the Kronecker product (e.g. Tucker, 1966, 283ff). We will not introduce special notation to distinguish between the two-way and three-way versions of X and G , as the appropriate version will be clear from the context. An alternative matrix representation is

$$X_k = AH_k B' + E_k \quad (k = 1, \dots, K) \quad (3)$$

where the H_k , the 'individual characteristic matrices' (Tucker, 1972), are equal to a linear combination of the R frontal slices, G_r , of the core matrix

$$H_k = \sum_{r=1}^R c_{kr} G_r \quad (4)$$

When instead of directly fitting the original data, cross-product, or covariance matrices are used as input, typically A and B will become identical or sign-permuted versions of each other, and the matrix C will in that case have a strong similarity to the compromise matrix in STATIS (Escoufier et al., this issue).

Tucker2 model

As indicated above the Tucker2 model contains only components

$$x_{ijk} = \sum_{p=1}^P \sum_{q=1}^Q a_{ip} b_{jq} h_{pqk} + e_{ijk}$$

$$(i = 1, \dots, I; j = 1, \dots, J; k = 1, \dots, K) \quad (5)$$

with the same definitions as above, except that $\mathbf{H} = (h_{pqk})$ is called the *extended core matrix*, because one way still has its original dimension, here K . A matrix formulation is identical to Equation (3), but the \mathbf{H}_k are unrestricted, compared to the Tucker3 model, where they have the form of Equation (4).

When instead directly fitting the original data, cross-product, covariance, or similarity matrices, generally \mathbf{A} and \mathbf{B} will become again identical or sign-permuted versions of each other. In this case the frontal core slices \mathbf{H}_k generally will be symmetric. The Tucker2 model is then identical to the IDIOSCAL model of Carroll and Chang (1970, 1972).

As hinted at by Harshman and Lundy (1984a,b), and worked out in Brouwer and Kroonenberg (1991b) a square extended core matrix may be optimally transformed to diagonality. If this is successful the model becomes identical to Harshman's (1970; Harshman and Lundy, this issue) PARAFAC model, or for symmetric frontal slices equal to the Carroll and Chang's (1970) INDSCAL model.

4. Algorithms

In this section we will discuss the algorithms that have been put forward for three-mode principal component analysis. In order to avoid repetition, we will primarily concentrate on algorithms for the Tucker3 model. Those for the Tucker2 model are essentially similar, in particular for the basic solution. In theory, any algorithm developed for the Tucker3 model can be used for the Tucker2 model by equating one of the component matrices to the appropriate unity matrix. For efficiency, we prefer to have a separate algorithm to solve the estimation of the Tucker2 model.

Tucker3 model

If we would compute all the components, thus $P = I$, $Q = J$, and $R = K$, then one could decompose most data matrices exactly into their components. However, in practical applications one is just interested in the two, three or four first components. This generally precludes finding an exact factorisation of \mathbf{X} into \mathbf{A} , \mathbf{B} , \mathbf{C} and \mathbf{G} . One, therefore, has to settle for an approximation, i.e., one has to find \mathbf{A} , \mathbf{B} , \mathbf{C} and \mathbf{G} such that the difference between the model and the data is minimal according to some loss function, or in slightly different terms, we have

to look for a best approximate factorization of the matrix X into A , B , C and G , according to the Tucker3 model.

In our case we define the loss function to be the least squares one, and propose to search for those A , B , C , and G such that

$$f(A, B, C, G) = \|X - AG(C' \otimes B')\|^2 \quad (6)$$

is minimal, where $\|\cdot\|$ denotes the Euclidean norm. The model is overidentified, because each of the component matrices is only determined upto a nonsingular transformation. To find a solution for f one has to place restrictions on the component matrices. It is convenient to carry out the minimisation under the restriction that A , B , and C are columnwise orthonormal (or suborthonormal), because this makes for an efficient and elegant algorithm. After estimates have been found, non-singular transformations can be applied to A , B and C without loss of generality, provided the core matrix G is multiplied with the corresponding inverse transformation matrices.

At present there are at least three algorithms with several variants to obtain estimates for the component matrices and the core matrix. The oldest one is due to Tucker (1966), but it has the disadvantage that the estimators have unclear properties. A first (alternating) least-squares algorithm was developed by Kroonenberg and De Leeuw (1980; see also Kroonenberg, 1983a Chapter 4, for a correction), using an eigenvalue-eigenvector algorithm by Bauer-Rutishauser for its inner iterations. Kroonenberg et al. (1989) showed that the algorithm could be slightly speeded up by replacing the Bauer-Rutishauser (BR) step by a Gram-Schmidt (GS) orthogonalisation. Kiers et al. (1992) showed how a really interesting increase in speed could be obtained by reorganising the computational process. Using regression techniques Weesie and Van Houwelingen (1983) developed a completely different algorithm especially designed to handle missing data. Kroonenberg (in preparation) adapted the Kroonenberg et al. (1989) algorithm to handle missing data by using an approach akin in spirit to the Expectation-Maximisation (EM) algorithm (Dempster et al., 1977), analogously to the procedure included in PARAFAC (Harshman and Lundy, this volume). All algorithms use loss function (6), and all complete data versions provide identical estimates, as do the missing data ones. From a user point of view the difference in algorithms is therefore not interesting, and need not be a concern. The TUCKALS programs contain the Kroonenberg et al. (1989), Kiers et al. (1992) and Kroonenberg (in preparation) algorithms, be it that at present only the first is commercially distributed.

In order to understand the results of the TUCKALS programs some basic understanding of the alternating least squares approach is necessary. The principle will be outlined using the Kroonenberg and De Leeuw approach to solving the estimation of the parameters in (6), and details of the other variants will be discussed in passing.

It can be shown that the least squares estimate for the core matrix G is

$$G = A'X(C \otimes B), \quad (7)$$

given that the component matrices are (columnwise) orthonormal. By substituting (7) into (6) f only depends on A , B , and C , and therefore it is sufficient to first estimate A , B , and C , and solve for G afterwards. The estimation proceeds as follows (α is the iteration counter).

TUCKALS3 ALGORITHM

- a. $\alpha = 0$.
- b. Initialise A_0 , B_0 , C_0 by using the Tucker (1966) approach.
- c. $\alpha = \alpha + 1$.
- d. A-substep:
Fix $B_{\alpha-1}$ and $C_{\alpha-1}$, and solve the least-squares problem for A with either BR or GS, to obtain a new A_α .
- e. B-substep:
Fix A_α and $C_{\alpha-1}$, and solve the least-squares problem for B with either BR or GS, to obtain a new B_α .
- f. C-substep:
Fix A_α and B_α , and solve the least-squares problem for C with either BR or GS, to obtain a new C_α .
- g. (Optional) Estimate missing data by using current estimates A_α , B_α , and C_α and the data X using model Equation (1).
- h. If the difference between successive iterations with respect to the loss function and the Euclidean norms of successive values of A , B , and C is not small enough return to c.

The major improvement by Kiers et al. (1992) is that the amount of multiplication involved in the step with the largest of I , J and K can be circumvented by cleverly rearranging the computations, so that manipulation with the original data matrix X is not necessary. For the estimation of missing data one has to include the extra step g., preventing the use of the Kiers et al. algorithm, exactly because it does not use the original data matrix. In Kroonenberg and De Leeuw (1980) the convergence properties of the basic algorithm were discussed. As in virtually all problems of this kind, only convergence to a local optimum is assured. Measures are taken to restart the algorithm in case of singularities due to very small components.

To initialise the algorithm A_0 , B_0 and C_0 are chosen in such a way they will solve Equation (6) exactly if such an exact solution is available. It can be shown that the eigenvectors associated with the largest eigenvalues of $U = X_{(I)}X'_{(I)}$ ($X_{(I)} \in \mathbb{R}^{I \times JK}$), $V = X_{(J)}X'_{(J)}$ ($X_{(J)} \in \mathbb{R}^{J \times KI}$), and $W = X_{(K)}X'_{(K)}$ ($X_{(K)} \in \mathbb{R}^{K \times IJ}$) will solve Equation (6) exactly if such a solution exists. These eigenvectors are, therefore, used to initialize the algorithm. This initial solution is, in fact, the Method I solution of Tucker (1966). Incidentally, this method was apparently independently discovered by Appellof and Davidson (1981).

Tucker2 model

As indicated above the estimation of the parameters for the Tucker2 model follows essentially along the same lines as that of the parameters in the Tucker3

model. The loss function f' may be written as

$$f'(A, B, H) = \|X - AHB'\|^2 = \sum_{k=1}^K \|X_k - AH_k B'\|^2 \quad (8)$$

with the same kind of definitions as above. When we eliminate the C-step (f.) from the TUCKALS3 algorithm, a parallel algorithm can be used for TUCKALS2, except that no components are specified for mode C. The model is not symmetric in its three modes as the Tucker3 model is, and therefore one has to make a decision which mode will remain uncondensed. This also means that not in all cases the acceleration due to Kiers et al. (1992) will lead to large gains in execution speed.

Sums-of-Squares notation

If we use \hat{z}_{ijk} for the implied data based on either model Equation (1) or Equation (5) the loss functions (6) and (8) may be written as

$$\sum_{i,j,k} e_{ijk}^2 = \sum_{i,j,k} (x_{ijk} - \hat{x}_{ijk})^2 = \sum_{i,j,k} x_{ijk}^2 - \sum_{i,j,k} \hat{x}_{ijk}^2 \quad (9)$$

which may be written in Sums-of-Squares notation as

$$SS(\text{Res}) = SS(\text{Tot}) - SS(\text{Fit}). \quad (10)$$

The quality of the fit of the overall solution can be evaluated by looking at the ratio $SS(\text{Fit})/SS(\text{Tot})$, which is the proportion sums-of-squares accounted for. When the raw scores have been centred in some way, and this is nearly always the case (see below), this ratio is equal to $R^2(\text{data, implied, data})$. It has been shown (Ten Berge et al., 1987), that when the algorithms have converged it is also true that

$$SS(\text{Res}_m) = SS(\text{Tot}_m) - SS(\text{Fit}_m), \quad (11)$$

where m stands for any level of any way of the data matrix. This is a powerful way to establish whether individual levels fit very well or very badly.

5. Input data and their manipulation

The TUCKALS programs are primarily geared towards metric or interval data, which are fully crossed with respect to all three ways. The missing data procedures allow some relaxation of the fully-crossing requirement, but the most stable results will be obtained without large numbers of systematically missing data. The programs do not have special ways of handling row-conditionality or matrix-conditionality. Such special structures should be taken care of beforehand, for example via centring and scaling (see below). Especially TUCKALS2 may be used for three-way two-mode data, such as multiple covariance matrices or (double-centred) (dis)similarity matrices (see Kroonenberg and Ten Berge,

1989, for details of the former case). In the latter case it is implicitly assumed that the dissimilarities are equal to squared distances rather than ordinary distances. If this is unacceptable, corrections should be made prior to the analysis. If the Tucker3 model is applied to sets of covariance matrices, the solution for the 'matrices' mode will be generally very similar to the comprise solution of STATIS (Escoufier et al., this issue). Unlike for the PARAFAC model (see Sands and Young's (1980) ALSCOMP3), there are (as of yet) no specific provisions in the programs for nonmetric data, such as optimal scaling or similar procedures for handling ordinal or nominal data (see Gifi, 1990, and Van der Burg, this issue, for information on optimal scaling), however one can analyse three-way interactions resulting from log-linear analysis or analysis of variance (see Kroonenberg, 1983a, Chapter 15, and Kroonenberg, 1989, for further details).

Generally, it is not advisable to analyse raw three-way data. As in two-way data, some kind of preprocessing in the form of subtracting certain means and equalizing scales of levels of modes is recommended to increase interpretability. Kroonenberg (1983a, Chapter 6), and especially Harshman and Lundy (1984b pp. 225-253) give detailed discussions of this problem. The most commonly applied centring are one or two fiber centring (removing row, column, or tube means; see Figure 2), and no, one, and very seldomly two size standardisations (equalizing the (mean) square in data slices; see Figure 1). Which and how many of these are necessary in any particular case is very much data dependent (see, however, Harshman and Lundy, 1984b, for a slightly different and more algebraic point of view).

Several centring can be performed by the programs, primarily on frontal slices of the data, but the programs are not specifically geared towards comprehensive data manipulation. In practice, the centring options suffice for most data sets, especially as by transposing the data matrix all desired centring can be performed, and centring on all three ways at the same time is hardly ever necessary. Full data manipulation can be performed with the separately available program NDIMIS3 (Brouwer and Kroonenberg, 1991a), or within Harshman and Lundy's PARAFAC (this issue). The latter program contains an (iterative) size-standardisation procedure for simultaneously size standardising two or three ways, and also has special procedures to handle sets of covariance and similarity matrices.

6. Programs

The versions of TUCKALS2 and TUCKALS3 discussed here are Version 5.0/PC for both programs. Around the time this paper appears, the TUCKALS PC-Interface should be available which will greatly facilitate the use of these programmes and NDIMIS3. However, the discussion here will be confined to the batch versions.

The primary output of TUCKALS2 (TUCKALS3) consists of the component matrices A , B (and C), and the core matrix H (G), as well as information on the convergence and fit of the overall solution. Apart from this, there is information about the input parameters, input data, and the like. To evaluate the quality of the solution several kinds of supplementary information can be requested, such as the fit per level of each way, residuals, fitted data, etc. To aid interpretation various kinds of line plots can be produced, such as pairwise plots of the components, plots of the fitted versus squared residuals, and joint plots of components from different ways. Several matrices containing such information can be written to external files for analysis with other programs. The TUCKALS programs also contain several transformation procedures, both for components and for core matrices. Moreover, options are included to compute combination-mode component scores and so-called core covariances. The details of the majority of this output will not be discussed here, but the reader is referred to the Manual of the programs (Kroonenberg and Brouwer, 1993), which also contains a list of publications using some of the more uncommon features.

Scaling of components and core matrix

The basic parameters of the Tucker3 model are the loadings for the three modes, A , B and C , and the elements of the core matrix G . There are several possibilities for scaling these basic parameters. The situation for the Tucker2 model is similar but will not be discussed explicitly.

Components of length one. In the algorithms, the component coefficient matrices are orthonormal i.e. they have orthogonal, length-one components. Therefore, the sizes of the coefficient vectors do not reflect the relative importance of the components, and the elements of the core matrix, g_{pqr} , directly reflect the size of the data. Furthermore, $\sum g_{pqr}^2 = \text{SS}(\text{Fit})$, each g_{pqr}^2 indicates the contribution of the (p, q, r) -combination component to the overall fit, and $g_{pqr}^2/\text{SS}(\text{Total})$ indicates the proportional contribution to the fit, or proportion explained variation (= sum of squares).

PARAFAC-scaled components. The disadvantage of the above scaling is that the absolute sizes of the coefficients are not comparable across modes, because of the generally different numbers of levels. As in PARAFAC (Harshman and Lundy, this issue) the coefficients can be made comparable by making the mean squared coefficients, rather than the lengths of the components, equal to one. Thus, for instance, the PARAFAC-scaled coefficients for the first mode would become $a_{ip}^* = a_{ip}(\sqrt{I})$. One could consider rescaling the core matrix with the inverse transformations, but this does not have any particular interpretational advantage.

Standard-PCA scaled components. In standard two-way component analysis the loadings are scaled such that their squared lengths reflect the size of the

eigenvalue of that component, and the eigenvalues add up to the number of levels. A similar scaling can be used in three-mode analysis, be it that coefficients scaled in this way are generally not correlations between the original entities in a mode and the component (see Harshman and Lundy, 1984a, p. 192ff. for a thorough discussion of this point). The scaling is thus such that $\bar{a}_{ip} = a_{ip} (\sqrt{\mu_p} I)$. Again one could scale the core matrix with the inverse transformations, but also this scaling does not seem to provide new interpretational insights.

External analysis. In certain applications component spaces are available from previous studies, and the question may arise whether this particular component space will also be applicable for a new set of data. An analysis in which such a component space is used as input, and kept fixed during the analysis, is called an *external analysis*. The programs include options to read in an external component space for each of the ways. A detailed example is presented in Van der Kloot and Kroonenberg (1985). External analysis may also be used for restarting insufficiently converged solutions.

Supplementary information. For a proper assessment of the fit of the three ways, it is necessary to have some insight into the structure of the *residuals* (= differences between the data and the implied data). A large residual sum of squares – $SS(\text{Residual}_m)$ – indicates that level m does not fit very well in the structure determined by the other levels. However, an extremely large residual sum of squares, often combined with a very large total sum of squares – $SS(\text{Total}_m)$ – is often indicative for some clerical error in the data. The size of a $SS(\text{Residual}_m)$ depends on its $SS(\text{Total}_m)$. Therefore, the *relative residual sum of squares* (= $SS(\text{Residual}_m)/SS(\text{Total}_m)$) should be used for the comparison of fit levels within a way. Often levels with large $SS(\text{Total}_m)$ s will fit better than those with small $SS(\text{Total}_m)$ s due to the least-squares procedures used. For each way the program provides a plot of the $SS(\text{Residual}_m)$ s versus the $SS(\text{Fit}_m)$ from which the relative performance of the levels can be gauged, i.e., both relative to each other, and relative to the overall fit/residual ratio. A more detailed analysis of the residuals is possible by investigating the $I \times J \times K$ block of residuals.

Joint Plots. Both in TUCKALS2 and TUCKALS3, it is very instructive to investigate the component coefficients of one mode (say, variables) jointly with those of another mode (say, conditions). This can be done by plotting them together in the same *joint plot*. For each core slice, say G_r in TUCKALS3 (and H_k in TUCKALS2), a joint plot for two component matrices, say A and B , can be constructed in such a way that the columns of A and B are close to each other. Closeness is measured as the sum of all $P \times Q$ squared distances $d^2(a_i, b_j)$, for all i and j . The construction is as follows. A G_r is decomposed via a singular value decomposition $G_r = U_r D_r V_r'$, and the orthonormal left-eigenvectors U_r , and the right eigenvectors V_r are combined with A and B respectively,

and the diagonal matrix D_r of singular values is divided between them in such a way that

$$A_r^* = (I/J)^{1/4} AU_r D_r^{1/2} \quad (12)$$

and

$$B_r^* = (J/I)^{1/4} BV_r D_r^{1/2}. \quad (13)$$

As $A_r^* B_r^{*'} = G_r B' = Y_r$, each element y_{ij}^r is equal to the inner product $a_i^* b_j^{*'}$, and provides the strength of the relationship between i and j in as far as it is contained in the r -th core slice. By simultaneously displaying the two modes in one plot, visual inferences can be made about their relationships. The joint plot is a close kin of Gabriel's (1971) biplot, and interpretational procedures developed by Gabriel (e.g. 1985) should be useful here as well. The construction for TUCKALS2 is obviously analogous.

Latent covariation matrix. In analyses of variables by conditions by subjects data the subject mode is often considered to be stochastic, rather than fixed. In that case a IJ by IJ multivariable/multicondition covariance matrix can be computed for the IJ variables-conditions combinations over all subjects. In an analogous manner the PQ by PQ 'covariation' matrix can be computed for the 'latent variables'-'prototype conditions' combinations over the 'idealized subjects' of the core matrix (see Tucker, 1966, or Kroonenberg, 1983a, Chapter 6, for such an interpretation of the components and the core matrix). The matrix is generally not a real variance-covariance matrix, except when the column means of the core matrix are zero, but only a sums-of-squares-and-cross-products matrix. In the TUCKALS3 case, its elements are the inner products

$$s_{pq,p'q'} = \sum_{r=1}^R g_{pq,r} g_{p'q',r}, \quad (14)$$

and in the TUCKALS2 case the summation is over k with h_{pqk} , rather than over r with g_{pqr} . The value of $s_{pq,p'q'}$ thus indicates the covariation of the pq -th and the $p'q'$ -th 'latent variables-prototype conditions' combination components. In Tucker's (1966) terminology, the core matrix is seen as a miniature of the original data set explaining most of its variance. In the same way one may say that the latent covariation matrix underlies the observed covariation matrix. For a more detailed description, one could consult Lohmöller (1978; see also Kroonenberg, 1983a, Chapter 13).

Component scores. In some applications it is useful to inspect the scores of all combinations of levels of two modes on the components of the third mode. For instance, for longitudinal data the scores of each subject-time combination on the variable components can be used to inspect the development of an individual's score on the latent variable or variable component over time. In some applications these component scores turn out to be a very successful summary of

the relationships involved. They serve as an intermediate level of condensation between the raw data and the three-mode model.

Within TUCKALS2 the component scores may be derived by rewriting the basic model Equation (5) for the Tucker2 model as follows

$$x_{ijk} = \sum_{p=1}^P a_{ip} d_{pjk} + e_{ijk}, \quad (15)$$

with

$$d_{pjk} = \sum_{q=1}^Q b_{jq} h_{pqk}. \quad (16)$$

A d_{pjk} can be thought of as the component score of individual k at occasion j on component p of the first mode A. By using Equation (4) one can define a similar expression for the Tucker3 model. Sometimes it is not very useful to inspect the plots of the scores of different components against one another, as is customary for component loadings. Instead, it is often more useful to inspect the component scores per component against their sequence numbers of the second or third mode. If one plots the component scores against each other, one obtains *trajectories* as are commonly presented in STATIS (Escoufier et al., this issue; for an example of this use see Kroonenberg, 1985).

Transformations. As mentioned above the Tucker2 and Tucker3 models are overidentified, and there is therefore no unique orientation of the axes, as in PARAFAC (Harshman and Lundy, this issue). The component matrices may be nonsingularly transformed, provided the core matrix is subjected to the inverse transformations. Alternatively, the core matrix may be transformed according to some criterion and the component matrices have to be adjusted accordingly. Both options are available in the programs: varimax and promax rotations of the (orthonormal) component matrices, and orthonormal and nonsingular transformations of the core matrices. The latter transformation is equivalent to investigating whether a PARAFAC solution exists for the model given the number of components. Details are contained in the Manual (Kroonenberg and Brouwer, 1993), as well as some methodological considerations with respect to choosing reasonable transformations.

7. TUCKALS3 application: Drunken twins

Data

Nick Martin and his colleagues of the Queensland Institute for Medical Research, Brisbane, Australia, collected a large amount of information from Australian twin pairs who received an acute or challenge dose of alcohol (see Martin et al., 1985, for the full experimental details). In this paper, we will

concentrate on 41 twins pairs who were measured at two separate occasions. At each occasion they were measured four times. The first time the subjects were sober. The other measurements were taken at hourly intervals after they had drunk 0.75g ethanol/kg body weight over a period of 20 minutes (which can make one fairly drunk, indeed). Here, we will only look at the variables: *Auditory Reaction Time* (ART), *Complex Reaction Time* (CRT), *Visual Reaction Time* (VRT), a speeded *Arithmetic Test* (ARI) consisting of simple addition and subtraction problems (number correct in two minutes; converted for this analysis into number of incorrect responses), and the subjects' judgements of their own *Drunkenness* (DRNK). The scores are coded in such a way, that high scores for all variables indicate a high influence of alcohol, i.e., long reaction times, large number of errors, and high ratings of intoxication.

In particular, we are dealing with a 82 (subjects) by 5 (variables) by 8 = 2*4 (measurement times) matrix. Before the three-mode analysis proper, the means of the variables at each measurement time were removed, and each variable was scaled over all measurements on that variable. The model used for this example is the Tucker3 model, in which components are computed for all three ways: 3 components for the subjects, 3 for the variables, and 2 for the measurement times.

Components

Variables. The structure in the three principal components of the variables has been enhanced by rotating them orthogonally according to a varimax criterion. The three axes V1, V2, and V3 can easily be labelled *Reaction Time* (RT), *Arithmetic* (ARI), and *Self-rated Drunkenness* (DRNK).

Time. The two time components are presented in a different fashion by plotting each component against time itself (see Figure 3). The time components get their full meaning in conjunction with the other modes, but it is evident, that the first component indicates the general *Persistence* of the effect of alcohol across time periods, and that the patterns of the first occasion (drawn lines) and second occasion (dashed lines) are very similar. There is good replicability, and there-

Table 1
Variable Components (after Varimax)

		Reaction Time	Arith- metic	Drunken ness
Auditory Reaction Time	ART	0.63	0.02	-0.08
Visual Reaction Time	VRT	0.57	-0.01	0.05
Complex Reaction Time	CRT	0.53	-0.01	0.02
Arithmetic Computation	ARI	-0.00	1.00	0.00
Self-rated Drunkenness	DRNK	0.02	0.00	0.99
Percentage Explained Variation		39	17	14

fore we will make no distinction between the two occasions. The same can be said with respect to the second component, which indicates the *Time-dependent reaction* of the subjects to the alcohol intake. In particular, the influence of alcohol is low at t_0 because the subjects were sober. At t_1 and t_2 the influence is most clearly felt, and falling off towards t_3 , three hours after the first consumption of alcohol. From the time components alone there is no telling which subjects on which variables follow the general pattern on which variables, for that we need the complete information from the analysis.

Subjects. The first two of the three components are shown in Figure 4. Without the labelling Figure 4 would show an amorphous cloud without any structure whatsoever. There are two ways to impart meaning to such clouds: via information present in the data set itself, i.e., in terms of its relationship with the components of the variables and those of the measurement times, and via external variables with additional information on the twins.

We have connected all twin pairs and labelled them according to type and sex. One would expect (1) that twins are closer together than randomly paired subjects, (2) that monozygote twins (connected with uninterrupted lines) are closer together than dyzygote twins irrespective of sex, and possibly (3) that dyzygote twins of the same sex (short dashed lines) are closer together than dyzygote twins of the opposite sex (long dashed lines).

To investigate the first hypothesis, Euclidean distances were computed between the twins using the three-dimensional subject space. These distances were compared with the distribution of distances computed for randomly connected pairs. Such pairs were created by randomly permuting the original subject coordinates. This procedure is called bootstrapping, and can be considered a permutation test (see e.g. Efron and Gong, 1983). In the present case 100 bootstrap samples were created and the average mean distance was computed over these hundred samples. The other two hypotheses were informally evaluated by comparing the mean distances.

The results, summarized in Table 2, show that, overall, twins are indeed closer together than randomly connected pairs. The observed mean distance is smaller than any bootstrap mean distance, and way beyond any reasonable confidence bounds. Looking at the twin types, various deviations can be observed from the general trend: female and mixed-sex dyzygotic twins are not very much below the bootstrap means, while the monozygotic twins clearly are, as are the male dyzygotic twins. Note, however, that type of twin is not related to any direction in the subject space.

Equally interesting is the simple separation by sex alone irrespective of twinship. We see clearly that alcohol has a different effect on males (δ) and females (η), and that we can associate a direction in the subject space with the sex difference. What this effect is, will become apparent from the investigation of the variable components, the time components, and the subject components together. Via a discriminant analysis of sex on the components, we have determined the direction which optimally separates men and women. In the

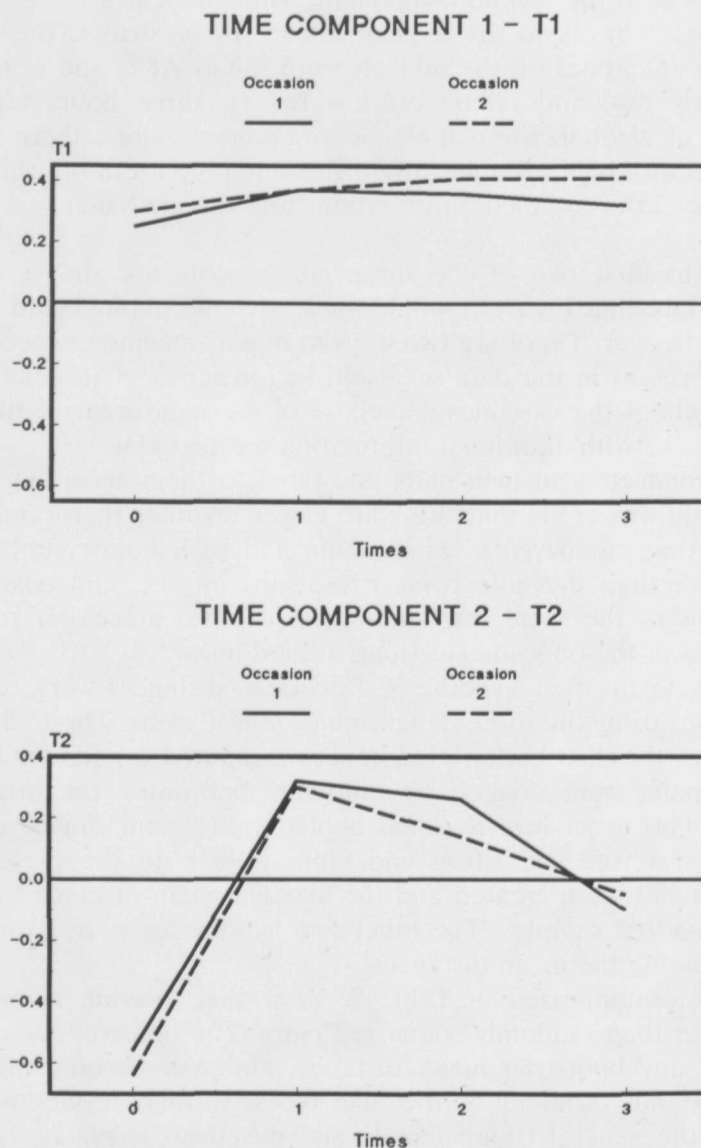


Fig. 3. Time components: T1 = Persistent effect of alcohol, T2 = Time dependent effect of alcohol.

sequel we will use the directions of the discriminant axes as new axes for the subjects, and continue to designate them S1, S2, and S3. In this way the first subject axis corresponds optimally with sex differences.

We have no further external information related to the separation between subjects. For instance, the scores on the subscales Extraversion, Psychoticism, Neuroticism, and Lie (or Social Desirability) from the Eysenck Personality

SUBJECT COMPONENTS

(labelled by sex and zygosity)

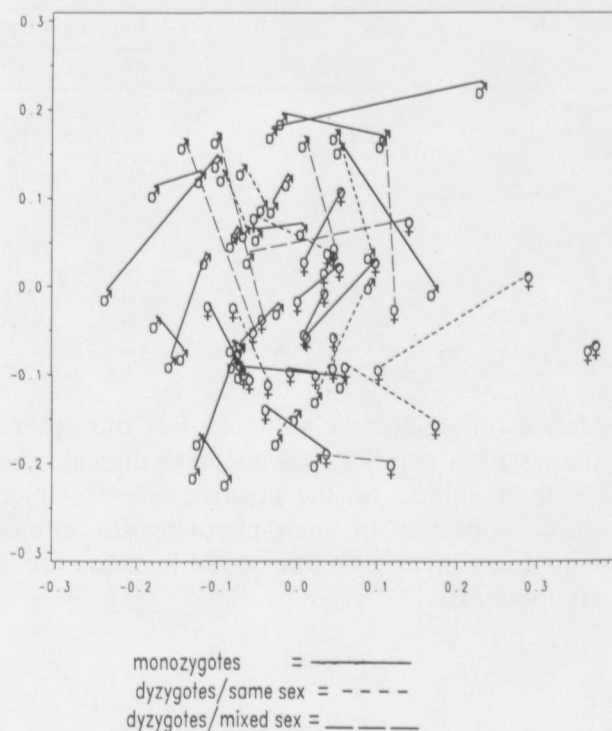


Fig. 4.

Questionnaire (Eysenck and Eysenck, 1975) did not show any relations with the subject components. For the present discussion, we will refer to a subject with a nonzero weight on one component and zero weights on all other components as a 'characteristic subject'. For the first component, this would mean that we have

Table 2
 Comparison of mean distances between twin pairs

Twin Type	Sex	N	Dist.	Bootstrap distances			
				Mean	S.D.	Min	Max
Monozygotes	Female	9	0.13	0.23	0.02	0.16	0.27
	Male	15	0.15	0.25	0.02	0.16	0.29
Dyzygotes	Female	6	0.21	0.25	0.03	0.17	0.29
	Male	5	0.10	0.23	0.04	0.15	0.31
	Mixed	6	0.21	0.25	0.03	0.16	0.32
Monozygotes	All	24	0.14	0.26	0.02	0.22	0.29
Dyzygotes	Single	11	0.16	0.24	0.02	0.18	0.29
	All	17	0.18	0.24	0.02	0.18	0.28
All twins	All	41	0.16	0.25	0.01	0.18	0.28

Table 3
Core Matrix

	S1 ♀ / ♂	S2	S3	Explained variability (%)		
				S1	S2	S3
<i>Persistent effect of alcohol (T1)</i>						
V1: Reaction Time	24	-26	-5	18	21	0
V2: Arithmetic	18	11	9	10	4	2
V3: Drunkenness	-2	-12	15	0	4	7
<i>Time-dependent effect of alcohol (T2)</i>						
V1: Reaction Time	-2	1	1	0	0	0
V2: Arithmetic	-3	-3	-0	0	0	0
V3: Drunkenness	-1	-6	7	0	1	1

a *Female* and a *Male* as characteristic subjects. For the other components we can only indicate them with a number plus a sign to indicate their location on a component, e.g. 2+ for a subject on the positive side of subject component 2. We will describe the properties of such characteristic subjects in terms of changes over time in their scores on the variable components, as expressed through the time components.

Core matrix

The relationships between the components of the various modes is contained in the core matrix as we pointed out before. For the present solution this core matrix is shown in Table 3. Note first of all that the T1 panel, referring to the persistent effect of alcohol on the subjects, has a rather complicated structure, suggesting that different characteristic subjects have quite different reactions towards alcohol.

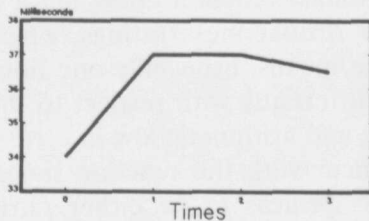
To explain this in detail we will have to look simultaneously at Table 3 and Figure 5, which shows for each time point the means of the variables averaged over replications, with reaction time also averaged over the three reaction-time measurements. The general conclusion from these figures is that reaction time stays at a higher level long after the alcohol intake, and long after the subjects say they feel less drunk. That the influence of alcohol is declining is borne out by the arithmetic test. The figures show what the scores are of the *Average Subject*. This Average Subject is located at the origin of the subject space, and will be the reference point for all further explanations.

With this background we can now turn to the investigation of the fundamental relationships in the data as contained in the core matrix. The first thing to note is that a strong temporal effect (over and above that already displayed in the average curves) is only evident in the self-ratings of drunkenness. The other variables or variable components only have the trend as portrayed in the first time component, viz. the persistent effect of alcohol. This means that the curves of the characteristic subjects especially for the after-alcohol periods stay parallel

MEAN VALUES

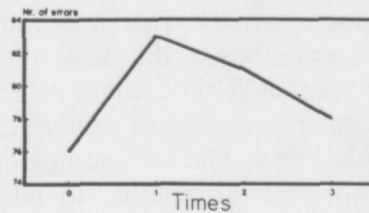
(=scores of Average Subject)

Reaction Time Pattern (RT)



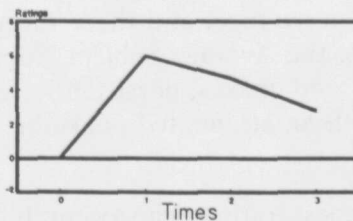
(Reaction time pattern is based on VRT, ART, and CRT scores)

Arithmetic (ARI)



(Number of errors are indirectly derived from numbers correct)

Self-rated Drunkenness (DRNK)



(Ratings based on both occasions)

Fig. 5.

to, either above or below, the mean curves. The easiest way to look at this core matrix is to describe the characteristic subjects one by one.

Characteristic Subject 1 (Female versus Male). Characteristic subject 1+ (Female) has persistently (T1) longer reaction times (V1) than average [core element (V1, S1, T1) = 24], also has persistently (T1) more arithmetic errors (V2) than average [core element (V2, S1, T1) = 18]. On the other hand, the characteristic subject 1- (Male) has persistently shorter reaction times than average, and persistently less arithmetic errors than average. Thus the general trend, as is embodied in the means, is elevated for females with respect to males for the performance variable components, while there is no appreciable sex-related deviation from the average in perceived drunkenness [core element (V3, S1, T1) = -2].

Characteristic Subject 2 (No specific relationship with external variables known). The 2+ subject shows persistently shorter reaction times than average [core element (V1, S2, T1) = -26], has persistently more arithmetic errors than

average [(V2, S2, T1) = 11]. He gives persistently lower drunkenness ratings [(V3, S2, T1) = -12]. Finally, also his time-dependent judgements of drunkenness are below average [(V3, S2, T2) = -6]. The time-dependent curve of Fig. 3 (inverted because of the minus sign) shows an inverse pattern to that of the average curve (Fig. 5), thereby attenuating the peak of the average curve. The 2 - subject shows the reverse pattern: persistently longer reaction times and less arithmetic errors. This is accompanied by higher drunkenness ratings, which tend to emphasize the peak already present in the means, especially one hour after alcohol. Thus alcohol affects these subjects differently with respect to the performance measures, either reaction time is long and arithmetic low in errors, or vice versa. The self-ratings of drunkenness concur with the reaction times, but not with arithmetic. In addition, the subjects profess to be either fairly sensitive to the alcohol (2 -), or are largely indifferent to it (2 +), as their time-independent curve counteracts the average one.

Characteristic subject 3 (No relationship with external variables known). Subject 3 + is about average on reaction time, but has persistently more errors and higher drunkenness than average, and these ratings are time-dependent in that they elevate the peak of the Average Subject. Subject 3 - is, of course, also average on reaction time, and makes persistently less errors and has lower ratings for drunkenness, with an attenuated peakedness directly after alcohol.

Summary. High drunkenness ratings can occur both with large number of errors, and with long reaction times. For some subjects their feeling of drunkenness is reflected in elevated scores for arithmetic, and not for reaction times, while for others it is the reverse, that is a high feeling of drunkenness is reflected in elevated scores for reaction time, but not for arithmetic. And when both performance measures are high or low the drunkenness ratings tend to be average. Furthermore, note that when there are differences between subjects on the drunkenness ratings higher than average scores tend to go together with higher peakedness, and conversely that low ratings go together with lower peakedness directly after alcohol. Thus emphasizing the sensitivity or insensitivity for alcohol.

Conclusion

By treating the example in some detail we have tried to convey some of the power of an integrated analysis of three-way data. In particular, we hope to have succeeded in showing that complex questions can be asked of complex data, but that such questions generally have complex answers. It demands a careful and thoughtful analysis, preferably with considerable theoretical insight into the subject matter.

In the present example, the theoretical background was not very profound,

but this is primarily due to the very common sense notions and variables in the research. The fact that our samples consisted of twin pairs does not seem to be very relevant for explaining differences in tolerance to alcohol. In that respect, sex does a far better job. However, it became evident that twins in general have more similar reactions than arbitrarily paired persons, be it that for dizygotes the situation is not unequivocal.

In addition to sex, one would like to find other external correlates to explain differences between subjects. Without such variables it is unrealistic to expect an understanding of differences between subjects on various measures. This becomes especially clear from the subject component for which we have external information. There we see a relative stronger deterioration of performance by women compared to men. It is interesting to see that this difference is not clearly related with differences in subjective perception of drunkenness by females and males.

8. Technical information

The programs were written in FORTRAN77, and the PC versions (Version 5.0) have been compiled with the Microsoft Fortran 5.1 compiler. The mainframe versions run satisfactorily on the IBM8083 and on a VAX under VMS. Previous versions have been installed on a large number of mainframes, but the present release is too young for such extensive testing.

The standard PC-versions need approximately 320K of free memory, but smaller or larger versions can be supplied upon request. The mainframe versions can be supplied with dynamic array allocation capabilities, which can run with a local dynamic array function. The variable array space depends on the size of the problem, especially the largest of I , J and K .

At present the input is based on old-fashioned fixed-column entry, but the TUCKALS PC Interface (written in Pascal 6.0) is under development (approximate release date: Spring 1993), so that the input can be entered directly on the screen. The input files can be saved and reused by the Interface so that running jobs with new parameters is only a matter of adapting the input screens. Eventually the Interface will allow a mixed interactive-batch approach towards running the programs. The intention is to include in the Interface options to run other three-way programs.

The TUCKALS programs and the preprocessing program NDIMIS3 are available from the author. For academic institutions, the present costs are f 300 (or approximately \$200) per program, but prices may change when the PC-Versions are fully operational.

Direct enquiries and orders should be sent to: P.M. Kroonenberg, Vakgroep Algemene Pedagogiek, Department of Education, Leiden University, Wassenaarseweg 52, 2333 AK Leiden. Tel. * - 31-71-273446/273434 (secr.)/273619 (fax). Electronic mail: "kroonenb@rulfsw.LeidenUniv.nl".

Acknowledgements

Many thanks go to Nick Martin of the Queensland Institute for Medical Research for making the data available. Piet Brouwer was the genius behind the Interface and he also kindly supplied the drawings for Figures 1 and 2, and Jos Henselmans assisted with the other drawings.

References

- Appelhof, C.J. and E.R. Davidson, Strategies for analyzing data from video fluorometric monitoring of liquid chromatographic effluents, *Analytical Chemistry*, **53** (1981) 2053–2056.
- Arabie, P., J.D. Carroll and W.S. DeSarbo, *Three-way Scaling and Clustering* (Sage, Beverly Hills, 1987).
- Basford, K.E., P.M. Kroonenberg, I.H. DeLacy and P.K. Lawrence, Multiattribute evaluation of regional cotton variety trials, *Theoretical and Applied Genetics*, **79** (1990) 225–234.
- Bentler, P.M. and S.-Y. Lee, Statistical aspects of a three-mode factor analysis model, *Psychometrika*, **43** (1978) 343–352.
- Bentler, P.M. and S.-Y. Lee, A statistical development of three-mode factor analysis, *British Journal of Mathematical and Statistical Psychology*, **32** (1979) 87–104.
- Bloxom, B., A note on invariance in three-mode factor analysis, *Psychometrika*, **33** (1968) 347–350.
- Brouwer, P. and P.M. Kroonenberg, User's Manual of NDMIS3. A program for manipulating three-way data, Technical report, Department of Education, Leiden University (1991a).
- Brouwer, P. and P.M. Kroonenberg, Some notes on the diagonalization of extended core matrices, *Journal of Classification*, **8** (1991b) 93–98.
- Carroll, J.D. and P. Arabie, Multidimensional scaling, *Annual Review of Psychology*, **31** (1980) 607–649.
- Carroll, J.D. and J.-J. Chang, Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart-Young' decomposition, *Psychometrika*, **35** (1970) 283–319.
- Carroll, J.D. and J.-J. Chang, IDIOSCAL: A generalization of INDSCAL allowing IDIOSyncratic reference systems as well as an analytic approximation to INDSCAL, Paper presented at the Spring Meeting of the Psychometric Society, Princeton, NJ, March 30–31 (1972).
- Coppi, R. and S. Bolasco (Eds.), *Multiway Data Analysis* (North-Holland, Amsterdam, 1989).
- Dempster, A.P., N.M. Laird and D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society B*, **39** (1977) 1–38.
- Efron, B. and G. Gong, A leisurely look at the bootstrap, the jackknife, and cross-validation, *The American Statistician*, **37** (1983) 36–48.
- Escoufier, Y., C. Lavit and P. Traissac, The ACT (STATIS method), *Computational Statistics & Data Analysis*, (1992).
- Eysenck, H.J. and S.B.G. Eysenck, *Manual of the Eysenck Personality Questionnaire* (Hodder & Stoughton, London, 1975).
- Gabriel, K.R., The biplot graphical display of matrices with applications to principal component analysis, *Biometrika*, **58** (1971) 453–467.
- Gabriel, K.R., Biplot display of multivariate matrices for inspection of data and diagnosis, in V. Barnett (Ed.), *Interpreting Multivariate Data* (Wiley, Chichester, UK, 1985) 147–174.
- Geladi, P., Analysis of multi-way (multi-mode) data, *Chemometrics and Intelligent Laboratory Systems*, **7** (1989) 11–30.
- Gifi, A., *Nonlinear Multivariate Analysis* (Wiley, Chichester, 1990).

- Harshman, R.A., Foundations of the PARAFAC procedure: Models and conditions for an 'explanatory' multi-modal factor analysis, *UCLA Working Papers in Phonetics*, **16** (1970) 1-84 [University Microfilms No. 10, 085].
- Harshman, R.A. and M.E. Lundy, The PARAFAC model for three-way factor analysis and multidimensional scaling, in: H.G. Law, C.W. Snyder Jr., J.A. Hattie, and R.P. McDonald (Eds.), *Research Methods for Multimode Data Analysis* (Praeger, New York, 1984a) 122-215.
- Harshman, R.A. and M.E. Lundy, Data preprocessing and the extended PARAFAC model, in: H.G. Law, C.W. Snyder Jr., J.A. Hattie, and R.P. McDonald (Eds.), *Research Methods for Multimode Data Analysis* (Praeger, New York, 1984b) 216-284.
- Harshman, R.A. and M.E. Lundy, PARAFAC: Parallel factor analysis, *Computational Statistics & Data Analysis* (1992).
- Israelsson, A., Three-way (or second order) component analysis, in: H. Wold and E. Lyttkens (Eds.), Nonlinear iterative partial least-squares (NIPALS) estimation procedures, *Bulletin of the International Statistical Institute*, **43** (1969) 29-51.
- Jennrich, R., A generalization of the multidimensional scaling model of Carroll & Chang, *UCLA Working Papers in Phonetics*, **22** (1973).
- Kiers, H.A.L., Comparison of 'Anglo-Saxon' and 'French' three-mode methods, *Statistique et Analyse des Données*, **13** (1988) 14-32.
- Kiers, H.A.L., Hierarchical relations among three-way methods, *Psychometrika*, **56** (1991) 449-470.
- Kiers, H.A.L., P.M. Kroonenberg and J.M.F. Ten Berge, An efficient algorithm for TUCKALS3 on data with large numbers of observation units, *Psychometrika*, **57** (1992) 415-422.
- Kroonenberg, P.M., *Three-mode Principal Component Analysis. Theory and Applications* (DSWO Press, Leiden, 1983a).
- Kroonenberg, P.M., Annotated bibliography of three-mode factor analysis, *British Journal of Mathematical and Statistical Psychology*, **36** (1983b) 81-113.
- Kroonenberg, P.M., Multivariate and longitudinal data on growing children. Solutions using a three-mode principal component analysis and some comparison results with other approaches, in: J. Janssen, F. Marcotorchino, and J.M. Proth (Eds.), *Data Analysis. The Ins and Outs of Solving Real Problems* (Plenum, New York, 1985) 89-112.
- Kroonenberg, P.M., Singular value decompositions of interactions in three-way contingency tables, in: R. Coppi and S. Bolasco (Eds.), *Multway Data Analysis* (North-Holland, Amsterdam, 1989) 169-184.
- Kroonenberg, P.M., Three-mode component models: A review of the literature, *Statistica Applicata. Italian Journal of Applied Statistics*, **4** (1992).
- Kroonenberg, P.M., Missing data in three-way analysis, (in preparation).
- Kroonenberg, P.M. and J. de Leeuw, TUCKALS2: A principal component analysis of three-mode data, Research Bulletin RB 001-77 (Department of Data Theory, University of Leiden, 1977).
- Kroonenberg, P.M. and J. de Leeuw, Principal component analysis of three-mode data by means of alternating least squares algorithms, *Psychometrika*, **45** (1980) 69-97.
- Kroonenberg, P.M. and J.M.F. ten Berge, Three-mode principal component analysis and perfect congruence analysis for sets of covariance matrices, *British Journal of Mathematical and Statistical Psychology*, **43** (1989) 63-80.
- Kroonenberg, P.M. and P. Brouwer, TUCKALS (version 5) User's manual, Technical Report (Department of Education, Leiden University, 1993).
- Kroonenberg, P.M., J.M.F. ten Berge, P. Brouwer and H.A.L. Kiers, Gram-Schmidt versus Bauer-Rutishauer in alternating least-squares algorithms for three-way data, *Computational Statistics Quarterly*, **4** (1989) 81-87.
- Law, H.G., C.W. Snyder Jr., J.A. Hattie and R.P. McDonald (Eds.), *Research Methods for Multimode Data Analysis* (Praeger, New York, 1984).
- Lee, S.-Y. and W.-K. Fong, A scale invariant model for three-mode factor analysis, *British Journal of Mathematical and Statistical Psychology*, **36** (1983) 217-223.
- Lohmöller, J.-B., How longitudinal factor stability, continuity, differentiation, and integration are

- protrayed into the core matrix of three-mode factor analysis, Paper presented at the European Meeting on Psychometrics and Mathematical Psychology, Uppsala, Sweden, June 16 (1978).
- Martin, N.G., J.G. Oakeshott, J.B. Gibson, G.A. Starmer, J. Perl and A.V. Wilks, A twin study of psychomotor and physiological responses to an acute dose of alcohol, *Behavior Genetics*, **15** (1985) 305-347.
- Sands, R. and F.W. Young, Component models for three-way data: ALSCOMP3, and alternating least squares algorithm with optimal scaling features, *Psychometrika*, **45** (1980) 39-67.
- Smilde, A.K., Three-way analyses: Problems and perspectives, *Chemometrics and Intelligent Laboratory Systems*, **10** (1992).
- ten Berge, J.M.F., J. de Leeuw and P.M. Kroonenberg, Some new results on principal component analysis of three-mode data by means of alternating least squares algorithms, *Psychometrika*, **52** (1987) 183-191.
- Tucker, L.R., Implications of factor analysis of three-way matrices for measurement of change, in: C.W. Harris (Ed.), *Problems in Measuring Change* (University of Wisconsin Press, Madison, 1963) 122-137.
- Tucker, L.R., The extension of factor analysis to three-dimensional matrices, in: H. Gullikson (Ed.), *Psychometric Theory* (Winston, New York, 1964) 110-19.
- Tucker, L.R., Some mathematical notes on three-mode factor analysis, *Psychometrika*, **31** (1966) 279-311.
- Tucker, L.R., Relationships between multidimensional scaling and three-mode factor analysis, *Psychometrika*, **37** (1972) 3-27.
- Tucker, L.R., Three-mode factor analysis applied to multidimensional scaling, Paper presented at the U.S.-Japan Seminar on Theory, Methods, and Applications of Multidimensional Scaling and Related Techniques, La Jolla, CA, August 20-24 (1975).
- van der Burg, E. OVERALS: Nonlinear canonical correlation with K sets of variables, *Computational Statistics & Data Analysis*, (1992).
- van der Kloot, W.A. and P.M. Kroonenberg, External analysis for three-mode principal component models, *Psychometrika*, **50** (1985) 479-494.
- Weesie, J. and J. van Houwelingen, *GEPCAM user's manual: Generalized Principal Components Analysis with Missing Values*, Technical report, Institute mathematical Statistics, University of Utrecht (1983).