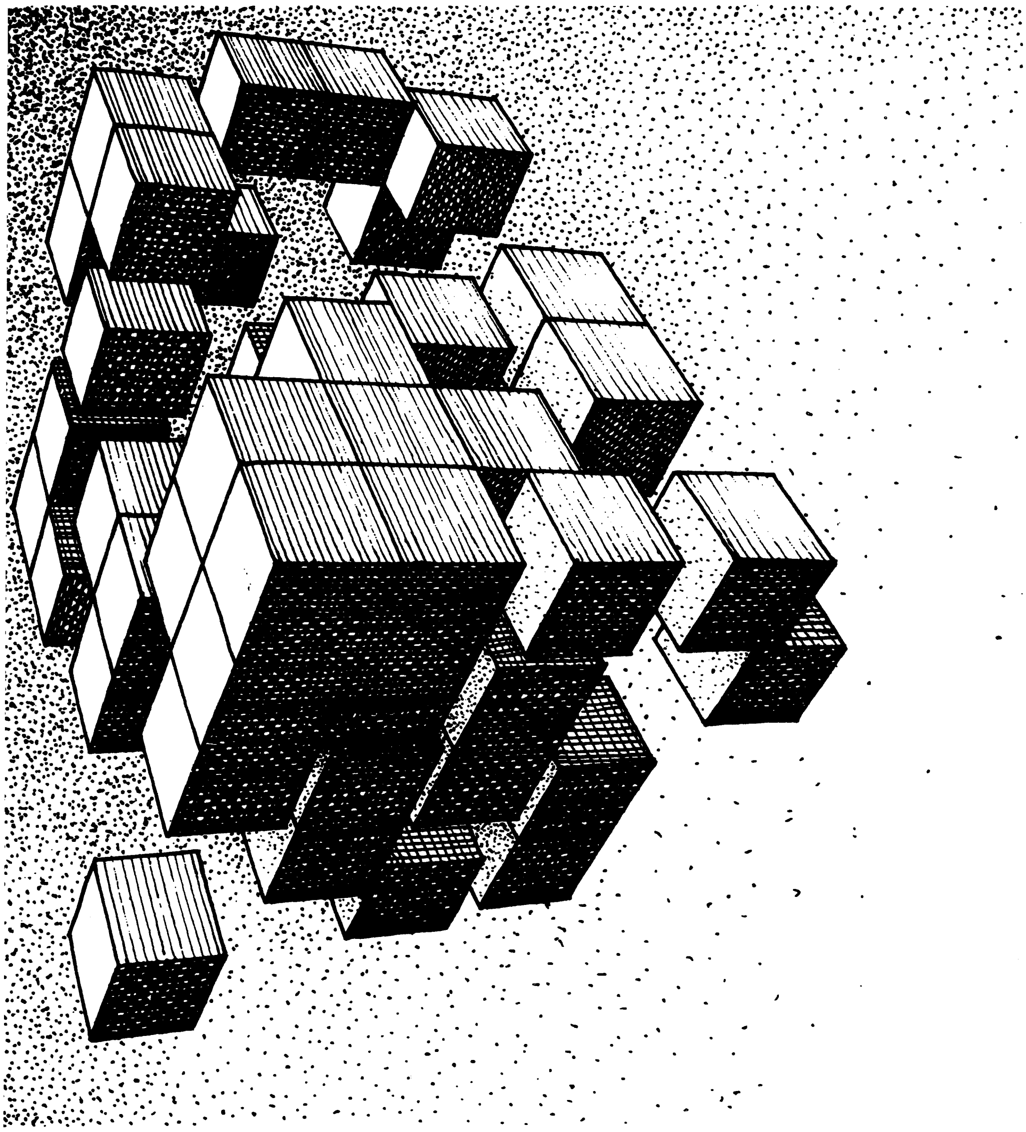
THREE-MODE
PRINCIPAL COMPONENT ANALYSIS

Pieter M. Kroonenberg



DSWO PRESS

THREE-MODE PRINCIPAL COMPONENT ANALYSIS
THEORY AND APPLICATIONS

M&T series volume 2

Other publications in this series:

Jacqueline Meulman, Homogeneity analysis of incomplete data.
M&T series 1, 1982

Pieter M. Kroonenberg, Tree-mode principle component analysis.
M&T series 2, 1983, reprint 1989

Jan de Leeuw, Canonical analysis of categorical data.
M&T series 3, 1984

Ronald A. Visser, Analysis of longitudinal data in behavioural and social research
M&T series 4, 1985

John P. van de Geer, Introduction to linear multivariate data analysis.
M&T series 5, volume 1 & 2, 1986

Jacqueline Meulman, A distance approach to nonlinear multivariate analysis.
M&T series 6, 1986

Jan de Leeuw, Willem Heiser, Jacqueline Meulman, Frank Critchley (editors), Multidimensional data analysis.
M&T series 7, 1987

Peter G.M. van der Heijden, Correspondence analysis of longitudinal categorical data.
M&T series 8, 1987

Jan van Rijckevorsel, The application of fuzzy coding and horseshoes in multiple correspondence analysis.
M&T series 9, 1987

Abby Israëls, Eigenvalue techniques for qualitative data.
M&T series 10, 1987

Eeke van der Burg, Nonlinear canonical correlation and some related techniques.
M&T series 11, 1988

Kees van Montfort, Estimating in structural models with non-normal distributed variables: some alternative approaches.
M&T series 12, 1989

Available from DSWO Press
Wassenaarseweg 52
2333 AK Leiden
The Netherlands
Tel. (071 - 273795)

THREE-MODE
PRINCIPAL COMPONENT ANALYSIS
THEORY AND APPLICATIONS

Pieter M. Kroonenberg

Department of Data Theory
Faculty of Social Sciences
University of Leiden

1983 DSWO Press, Leiden
Reprint 1989

CIP-DATA KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Kroonenberg, Pieter M.

Three-mode principal component analysis: theory and applications / Pieter M Kroonenberg. - Leiden: DSWO Press. - Ill. - (M&T series; vol. 2)
Ook verschenen als proefschrift Leiden. - Met index, lit. opg. - Met samenvatting in het Nederlands.
ISBN 90-6695-002-1
SISO 300.6 UDC 301.06
Trefw.: sociaal-wetenschappelijk onderzoek / statistiek / data-analyse.

Cover design and Prelims Marjorie Meulman
Cover drawing Siep Kroonenberg
Printed by 'Reprodienst, Subfaculteit Psychologie, Rijksuniversiteit Leiden'
and by Printing office 'Karstens drukkers bv, Leiden'

ISBN 90-6695-002-1

"...statisticians have no fields of their own from which to harvest their data. Statisticians get all their data from other fields, and from all other fields, wherever data are gathered ...
Ours is a symbiotic way of life, a marginal and hyphenated existence. We resemble the professional harvesters of wheat and grains on [the] Great Plains, who own no field of their own, but harvest field after field, in state after state, and lead a useful, rewarding, and interesting life - as we do."

Leslie Kish, *JASA*, 1978, 73, 1

PREFACE

The purpose of this book is three-fold. In the first place it is a monograph on three-mode principal component analysis. An attempt has been made to discuss virtually all issues in connection with the technique, and to collect and evaluate the literature on the subject which has appeared since its introduction in the sixties by Ledyard R. Tucker at the University of Illinois.

Secondly, this book introduces improved estimation procedures of the parameter in three-mode principal component models, and treats a number of consequences of these new procedures. Furthermore, some theoretical contributions with respect to transformations of core matrices are presented.

Thirdly, this book aims to provide a guide for social scientists and others who wish to apply three-mode analysis to their data. General issues with regard to what goes in and what comes out of a three-mode analysis are discussed, while some thirteen data sets from varying backgrounds and composition are analysed. This book also bring together almost the entire applied literature as part of the references, thus allowing researchers to compare their results with others from their own field of interest. A classification of these applied papers has been included in the Appendix.

Many people assited in the preparation of this book. The core of this book and many technical details were conceived by Jan de Leeuw. Leo van der Kamp, Wim van der Kloot, and Charles Lewis read various part of the book, suggested improvements, corrected errors, and helped clarify many ideas expressed here. Frits Goossens, Ineke Stoop, Jan Swaan, Albert Verbeek, Ron Visser, and Rien van IJzendoorn assisted with parts of the text. Jeanet Bus, Peter van der Heijden, Wim van der Kloot, Cor Lammers, Jan de Leeuw, and Tom van der Voort provided or suggested data sets for inclusion. Ineke Smit polished the English as far as was possible without specific knowledge of the subject matter; Siep Kroonenberg made the cover draw-

ing; Cora Jongsma expertly typed the text, and patiently included the many formulas and revisions; Piet Brouwer joined in the drudgery of checking the manuscript; Lutgart Balfoort assisted in getting this book into print.

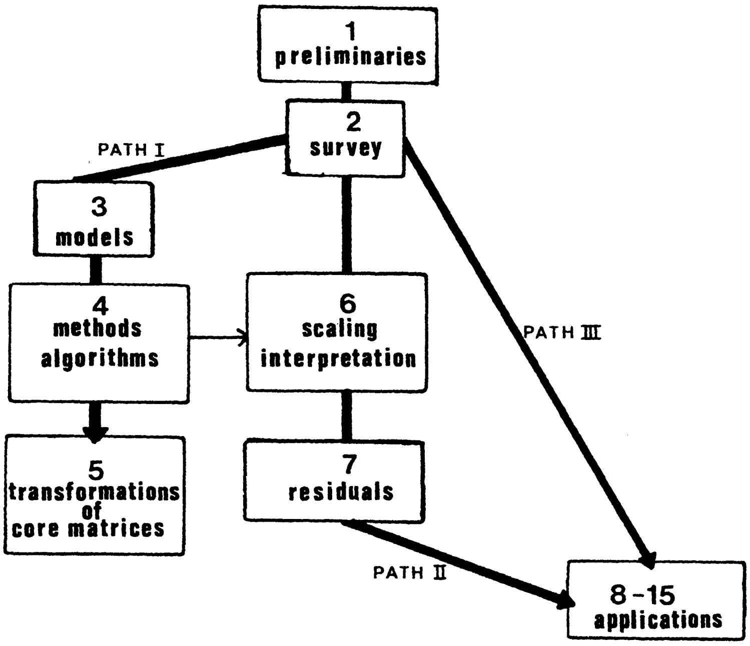
Leiden, January 1983

CONTENTS

DETAILED CONTENTS

PRELIMINARIES

1



1.1 INTRODUCTION

Investigating the relationships between variables is a favourite research activity of social scientists. They often want to explore the structure of a large body of data. To understand this organization the data have to be condensed in one way or another, and the raw data have to be combined to form summary measures which are more easily comprehended.

Among the most popular methods to achieve such condensation and summarization are principal component analysis and multidimensional scaling. Different variants exist in both cases, and their appropriateness varies with the research design. 'Standard' principal component analysis is applied when observations are available for a number of variables, and it is desired to condense these variables to a smaller amount of independent 'latent' variables or components. Similarly, 'standard' multidimensional scaling is applied when similarity measures are available for a number of variables, in this context usually called stimuli, and insight is desired into their structural organisation.

In many research designs, observations on variables have been made under a number of conditions, or at various points in time, or similarity measures have been produced by a number of persons, etc. In such cases, where the data can be classified by three kinds of quantities, or modes, e.g. subjects, variables, and conditions, the standard variants no longer suffice.

The extra mode in the design requires an extension of those standard techniques. It is, of course, possible to mould the data into the standard two-mode (or two-way) format by rearranging them into two-mode matrices, but this entails losing a part of the information which could be very important for the understanding of the organization of the data as a whole.

Since the introduction of three-mode principal component analysis by Tucker in 1964, and of individual differences scaling by Bloxom (1968b), Carroll & Chang (1970), and Horan (1969), considerable progress has been made in finding ways to confront the summarization and condensation of three-mode data. This has mainly been done by adapting the standard techniques to make them fit the problems created by the extra mode. That this introduces many complications will become clear in the sequel.

1.2 SOME EXAMPLES

To get a general idea of the kind of research problems three-mode principal component analysis can handle, we first give some examples of typical applications.

Semantic differential data. A classical example of three-way classified data can be found in the work of Osgood and associates (e.g. Osgood, Suci & Tannenbaum, 1957). In the development and application of semantic differential scaling, subjects have to judge various concepts using bi-polar scales of adjectives. Such data used to be analysed averaged over subjects, but the advent of three-mode principal component analysis and similar techniques has made it possible to analyse the subject mode as well in order to detect individual differences with regard to the semantic organization of the relations between scales and concepts. An example of such a study can be found in Snyder & Wiggins (1970). In Chapter 9 we present an example of this type of data.

Similarity data. Three-way similarity data consisting of stimuli x stimuli x subjects are generally analysed with individual differences scaling programs, such as INDSCAL (Carroll & Chang, 1970) and ALSCAL (Takane, Young, & De Leeuw, 1977). However, when the data are asymmetric and/or a more general model is required, three-mode principal component analysis can provide useful insight. See Chapter 3 for details on individual differences scaling and its relation with three-mode component analysis, and Chapter 11 for an empirical comparison of the techniques on the same data.

Asymmetric similarity data. Van der Kloot & Van den Boogaard (1978) collected data from 60 subjects who rated 31 stimulus persons on 11 personality trait scales. In the original report the data, which can be considered asymmetric similarity data, were first averaged over subjects, and subsequently analysed by canonical discriminant analysis using the stimulus persons as groups. Van der Kloot & Kroonenberg (1982) used three-mode principal component analysis on the original data to assess the individual differences and the extent to which the subjects shared the common stimulus and scale configurations. A summary of the results can be found in Chapter 10. The example in Chapter 2 on similarities between Dutch political parties also falls into this class of applications.

Multivariate longitudinal data. In the social sciences, multivariate longitudinal data pose problems for many standard techniques. (See Visser (1982) for a detailed review of techniques useful for such data in psychology). There are often too few observations and/or too many points in time for the 'structural approach' to the analysis of covariance matrices (Jöreskog & Sörbom, 1977), or too few points in time and/or too many variables for multivariate analysis of time series by some kind of ARIMA model (see e.g. Glass, Wilson & Gottman, 1975, or Cook & Campbell, 1979, Ch. 6). In such situations three-mode principal component analysis can be very useful, especially for exploratory purposes.

Lammers (1974) presented an example of longitudinal data with a relatively large number of variables (22), and only a limited number of points in time (11 years). The aim of the study was to determine whether some of the 188 hospitals measured showed different growth patterns or growth rates compared to the other hospitals. A re-analysis of these data is presented in Chapter 13.

In Chapter 14 we present a three-mode analysis of typical learning data collected by Bus (1982). In this case there were only six observational units (children), who had scores on five tests, but measures were available for 37 more or less consecutive weeks.

Three-way contingency tables. One of the ways to study interactions in large two-way contingency tables is by correspondence analysis. In its most common form, this technique is an analysis of the dependencies of the column and row categories of a contingency table by means of a so-called 'singular value decomposition' (see section 2.2) of the standardized residuals. A similar procedure can be defined for three-way tables using three-mode principal component analysis instead of the singular value decomposition. This approach is outlined and illustrated in Chapter 15 with data from three different elections for the election wards or precincts of Leiden.

1.3 ORGANIZATION OF THIS BOOK

The aim of this book is to treat three-mode principal component analysis with all its possibilities and limitations. We will pay attention to both theoretical and practical aspects of the technique, and therefore the level of the exposition will vary in mathematical sophistication. The theory is mainly dealt with in Chapters 3, 4, and 5, and the applications in Chapters 8 through 15. Chapter 2 provides a quick run-through of the entire book, and Chapters 6 and 7 are intermediary in the sense that they treat the theory necessary for a detailed understanding of the technical aspects of the applications and their interpretation.

The reader only interested in the technical aspects of three-mode principal component analysis, (or *three-mode analysis* for short), and its relation to other models and methods of analysis, should follow PATH I in Fig. 1.1, reading only Chapters 1 (preliminaries), 2 (survey), 3 (models), 4 (methods and algorithms), and 5 (transformations of core matrices). In Chapter 6 (scaling and interpretation) and 7 (residuals) some further technical information on matters that precede and follow a three-mode analysis can be found.

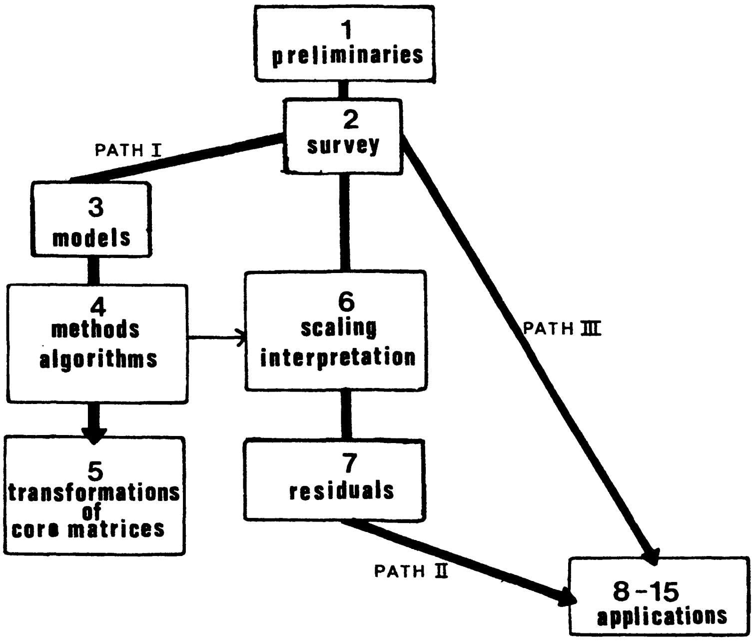


Fig. 1.1 *Organization of this book*

A reader only interested in the scope of the methods, and whose interest does not go beyond a basic notion of what three-mode analysis is about, should follow PATH III and read Chapters 1, 2, and the applications. However, in order to take full advantage of three-mode principal component analysis in practical situations it is best to read Chapters 6 and 7 as well, i.e. following PATH II, as in Chapter 6 the scaling of the input is discussed, as well as some interpretational aspects of the output which are helpful in understanding the peculiarities of the data at hand, and in Chapter 7 methods are described to assess the quality of the solutions obtained.

In order to facilitate the reading of single chapters, a three-mode *glossary* of the major terms used in this book has been included as next section.

For people looking for specific applications within their field of interest a large number of pertinent papers have been included in the references. To improve the usefulness of their inclusion these papers have been classified according to subject matter and data type. A list of papers referring to computer programs has been included as well.

1.4 THREE-MODE GLOSSARY

Basic terms

combination-mode (ij)	- Cartesian product of two (elementary) modes i and j; "i outer loop, j inner loop"; see Tucker (1966a, p. 281)
combination-mode matrix	- two-mode matrix with one (elementary) mode (usually columns) and one combination-mode (usually rows).
core matrix	- three-mode matrix, which contains the relations between the components of the various modes; its size is usually $s\times t\times u$, where s, t, and u are the number of components for the first, second, and third mode respectively.
element (of a mode)	- generic term for a variable (subject, condition, etc.) in a mode.
extended core matrix	- three-mode core matrix, of which one of the dimensions is equal to the number of elements in that mode; its size is usually $s\times t\times n$, where n is the number of elements in the third mode.
frontal plane	- $s\times t$-slice of an (extended) core matrix, or $\ell\times m$-slice of a three-mode data matrix.

mode (or elementary mode); way	- collection of indices by which the data can be classified; way and mode are here used as synonyms; for a different usage of the word 'mode' in the same context see Carroll & Arabie (1980).
reduced mode	- mode of which principal components have been computed.
three-mode matrix (-array)	- collection of numbers which can be classified in three (different) ways. i.e. using three indices; the numbers can thus be arranged in a three-dimensional block.

Methods

Covariance structure approach	- In this method the subject mode is treated as a random variable, and the analysis is performed on the combination-mode covariance matrix of the other two modes. Solutions can be obtained by maximum likelihood estimation, or generalized least squares procedures. An a priori structure for the component matrix and the core matrix can be specified.
Alternating Least Squares (ALS)	- An iterative method to solve large and complex models by breaking up the total number of parameters in a number of groups, each of which can be estimated conditional on the fixed values of the parameters in the other groups.
Partial Least Squares (PLS)	- See Alternating Least Squares

Tucker's (1966a) Method I - Standard principal component analysis on each of the three combination-mode matrices, and subsequent combination of the three solutions to form the core matrix.

Tucker's (1966a) Method II - Standard principal component analysis on two combination-mode matrices, combined with a clever juggling to compute an approximate core matrix and the third principal component matrix without resorting to solving the eigenvalue - eigenvector problem for the largest mode. Appropriate for data sets with one very large mode, usually individuals.

Tucker's (1966a) Method III- Method to analyse multitrait-multimethodlike covariance and correlation matrices. Forerunner of the covariance structure approach.

Models

CANDECOMP - Carroll & Chang (1970). T3 with a three-way identity matrix as core matrix, or equivalent to INDSCAL with different reduced modes.

IDIOSCAL - Carroll & Chang (1972). As T2, but the two reduced modes are equal, and thus the extended core matrix is symmetric in its frontal planes. Allows for both idiosyncratic rotations of axes in the common stimulus space, and individually different weighting of these axes. Component matrices are not necessarily orthogonal.

INDSCAL	- Carroll & Chang (1970). As IDIOSCAL, but with the additional restriction, that the frontal planes are diagonal, i.e. no idiosyncratic rotations are allowed. The model can also be interpreted as having three reduced modes of equal numbers of components, and a three-mode identity core matrix.
PARAFAC	- Harshman (1970, 1972a,b, 1976). Parallel profiles factor analysis. PARAFAC1 is equal to CANDECOMP. PARAFAC2 is similar to IDIOSCAL, but it specifies a common weighting of the axes of the stimulus space. However, idiosyncratic rotations of these axes are allowed.
Three-mode Scaling	- Tucker (1972a). As the Tucker3 model, but two of three reduced modes are equal. Core matrix has symmetric frontal planes.
Tucker2 model (T2)	- Israelsson (1969). Model specifies two unequal reduced modes with an unrestricted extended core matrix.
Tucker3 model (T3)	- Tucker (1966a). Three unequal reduced modes with an unrestricted core matrix.
Tucker's common factor model	- Tucker (1966a). As T3, but unique variances are specified for the combination-mode covariance matrix.

Terms with special definitions

component	- vector of loadings (e.g. g_p, h_q, e_r)
component weight	- eigenvalue, indicating the amount of variation explained by the component corresponding to the eigenvalue.

i-mode	-	first mode
j-mode	-	second mode
k-mode	-	third mode
SS(Fit)	-	sum of squares of the estimated data values, derived from the fitted model
SS(Res)	-	residual sum of squares
SS(Tot)	-	total sum of squares of the data
standardized component weight	-	component weight divided by the total sum of squares of the data
standardized sum of squares (St.SS)	-	sum of squares divided by the total sum of squares of the data.
variation	-	general term to indicate the sum of squares, generally of data values; depending on their scaling variations may be sums of squares, average sums of squares, or variances.

1.5 NOTATION

Matrices

$R^{a\times b}$	set of real matrices with a rows and b columns
$[[X]]$, $X \varepsilon R^{a\times b}$	$\surd(\sum_{i=1}^{a} \sum_{j=1}^{b} x_{ij})$; Euclidean norm
tr X, $X \varepsilon R^{a\times a}$	$\sum_{i=1}^{a} x_{ii}$; trace of X
$C = X \otimes Y$, $X \varepsilon R^{a\times b}$, $Y \varepsilon R^{c\times d}$	$c_{mn} = x_{ij} y_{k\ell}$; m=1,...,ac; n=1,...,bd Kronecker product
diag (X)	$d_{ij} = x_{ij}$ if i=j, and 0 otherwise

Data

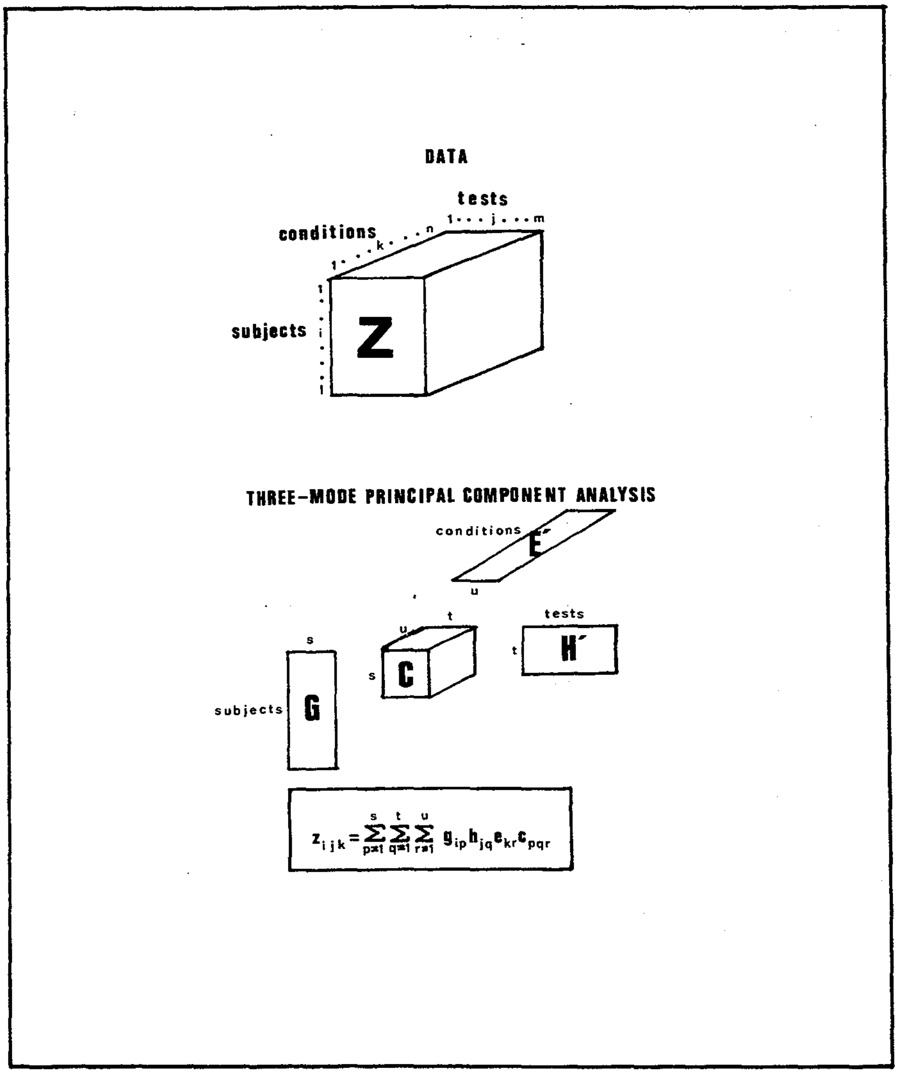
$\mathrm{Z} = \{\mathrm{z}_{ijk}\}$	three-mode data matrix; $i=1,\ldots,\ell$ (rows), $j=1,\ldots,m$ (columns), $k=1,\ldots,n$ (frontal planes)
$\hat{\mathrm{Z}} = \{\hat{\mathrm{z}}_{ijk}\}$	three-mode matrix with data values estimated from the fitted model

Model

first mode	second mode	third mode	
ℓ	m	n	number of elements
i	j	k	index of elements
s	t	u	number of components ($\ell \geq s$; $m \geq t$; $n \geq u$)
p	q	r	index of components
$\mathrm{G} = \{\mathrm{g}_{is}\}$	$\mathrm{H} = \{\mathrm{h}_{jq}\}$	$\mathrm{E} = \{\mathrm{e}_{kr}\}$	component matrix (orthonormal for Tucker models); for definition see *Theorem 4.1.*
$G = \{g_{is}\}$	$H = \{h_{jq}\}$	$E = \{e_{kr}\}$	component matrix for Tucker Methods for definition see *Theorem 4.2.*
g_p, g_p	h_q, h_q	e_r, e_r	components, vectors of loadings
λ_p	μ_q	ν_r	standardized component weights (eigenvalues of P (or *P*), Q (or *Q*), R (or *R*), respectively [for definition see *Theorem 4.1* (or *Theorem 4.2*)]
$\mathrm{C} = \{\mathrm{c}_{pqr}\}$			core matrix of Tucker3 model
$C = \{c_{pqr}\}$			core matrix of Tucker3 model for Tucker Methods
$\tilde{\mathrm{C}} = \{\tilde{\mathrm{c}}_{pqr}\}$			extended core matrix of Tucker2 model
$\tilde{C} = \{\tilde{c}_{pqr}\}$			extended core matrix for Tucker Methods

SURVEY

2



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

2.1 INTRODUCTION

In this chapter we first present the three-mode principal component model on a conceptual level by providing various informal descriptions of the model. Secondly, an outline of some technical aspects connected with analysing this type of model will be presented, and finally an example is used to illustrate some of the major aspects and possibilities of analysing three-mode data with the three-mode principal component model. In this way the chapter will serve as an introduction to the subject matter and terminology of this book.

The chapter aims to be comprehensible for the relatively uninitiated, but a basic working knowledge of standard principal component analysis is essential, as is insight into eigenvalue-eigenvector problems.

2.2 INFORMAL DESCRIPTIONS

In this section we will present three more or less different ways of looking at three-mode principal component analysis: we start with questions a researcher might ask about three-mode data, and discuss the way these questions fit into the framework of a three-mode principal component model. Next, we take a structural point of view, i.e. postulate some structural relationships, and investigate how real data might be described by a combination of structural parameters. Finally, we will take a methodological point of view, and demonstrate how three-mode principal component analysis is a generalization of standard principal component analysis and so-called singular value decomposition.

Research questions arising from three-mode data. After collecting information from a number of subjects on a large number of variables, one often wants to know whether the observed scores could be described as combinations of a smaller number of more basic variables or so-called *latent variables*.

As an example one could imagine that the scores on a set of variables are largely determined by linear combinations of such latent variables as the arithmetic and verbal content. The latent variables - arithmetic and verbal content - can be found by a standard principal component analysis.

Suppose now in the same example that the researcher has administered the variables a number of times under various conditions of stress and time limitations. The data are now classified by three different types of quantities or *modes* of the data: subjects, variables, and conditions. Again the researcher is interested in (1) the components of the variables which explain the larger part of the variation in the data. Moreover, he wants to know (2) if general characteristics can be defined for subjects as well. To put it differently, he wants to know if it is possible to see the subjects as linear combinations of 'idealized subjects'. In the example we could suppose that the subjects are linear combinations of an exclusively mathematically gifted person and an exclusively verbally gifted person. Such persons are clearly 'ideal' types. Finally a similar question could arise with respect to conditions: (3) can the conditions be characterized by a set of 'idealized' or 'prototype' conditions?

Each of the three questions can be answered by performing principal component analyses for each mode. In fact, the same variation present in the data is analysed in three different ways. Therefore, the components extracted must in some way be related. The question is how? In order to avoid confusion in answering this question, we will call the variable components *latent variables*, the subject components *idealized subjects* and the condition components *prototype conditions* (see section 6.8 for a discussion of designating components in this manner).

With respect to the relation between the components of the three modes one could ask questions like: "Do idealized subject 1 and idealized subject 2 react differently to latent variable 2 in prototype condition 1?" Or one could ask: "Is the relation between the idealized subjects and the latent variables different under the various prototype conditions?" By performing three separate component analyses such questions are not immediately answerable, as one does not know how to relate the various components. The three-mode principal component model, however, explicitly specifies how the relations between the components can be determined. The three-mode matrix which embodies these relations is called the *core matrix* as it is assumed to contain the essential characteristics of the data.

Structure: raw scores derived from idealized elements. It is often useful to look at three-mode principal components starting from the other end, i.e. starting with the core matrix. For example, we pretend to know how an exclusively mathematically gifted person scores on a latent variable which has only a mathematical content, and on a latent variable which has solely verbal content. We pretend to know these scores under a variety of prototype conditions. In other words, we pretend to know how idealized subjects react to latent variables under prototype conditions. As in reality we deal with real subjects, variables and conditions, we have to find some way to construct the actual from the idealized world. A reasonable way to do this is to suppose that a real subject is a mixture of the idealized individuals, and make an analoguous assumption for variables and conditions, so that the real scores can be thought of as combinations of mixtures of idealized entities.

What is still lacking is some rule which indicates how the idealized quantities can be combined into real values. One of the simplest ways to do this is to weight each, for instance, latent variable according to its average contribution over all subjects and conditions, and add the weighted contributions. In more technical terms, each real variable is a linear combination of the latent variables.

We will show how we may construct the score of an indivudual i on a test j under condition k from known idealized quantities. Suppose we have at our disposal 2 idealized persons (p_1,p_2), 2 latent variables (q_1,q_2), and 2 prototype conditions (r_1,r_2), and we know the scores of

subject p_1 on variable q_1 under condition r_1 : $c_{p_1q_1r_1}$ or c_{111};
subject p_1 on variable q_1 under condition r_2 : $c_{p_1q_1r_2}$ or c_{112};
subject p_1 on variable q_2 under condition r_1 : $c_{p_1q_2r_1}$ or c_{121};
subject p_1 on variable q_2 under condition r_2 : $c_{p_1q_2r_2}$ or c_{122}.

Similarly we know the scores of subject p_2: c_{211}, c_{212}, c_{221}, and c_{222}. In other words, we know all the elements of the core matrix. As mentioned above we want to construct the score of real subject i, on a real variable j, under a real condition k. We will do this sequentially, and assemble all the parts at the end. We start with the assumption that the score of a real subject i on the latent variable q_1 under a prototype condition r_1 is a linear combination of the scores of the idealized persons p_1 and p_2, using weights g_{ip_1} and g_{ip_2}:

$$s_{iq_1r_1} = g_{ip_1}c_{p_1q_1r_1} + g_{ip_2}c_{p_2q_1r_1}$$

$$\text{or} \quad s_{i11} = g_{i1}c_{111} + g_{i2}c_{211}.$$

Similarly, for variable q_2 under condition r_1:

$$s_{i21} = g_{ip_1}c_{p_1q_2r_1} + g_{ip_2}c_{p_2q_2r_1}$$

$$= g_{i1}c_{121} + g_{i2}c_{221} ,$$

and the other variable-condition combinations:

$$s_{i12} = g_{i1}c_{112} + g_{i2}c_{212},$$
$$s_{i22} = g_{i1}c_{122} + g_{i2}c_{222}.$$

The weights g_{i1} and g_{i2} thus indicate to what extent the idealized subjects p_1 and p_2 determine the real subject i. The assumption in this approach is that these g_{i1} and g_{i2} are independent of the test and the condition under which the subject is measured. All interactions between subjects, variables, and conditions arise from interactions between the idealized entities, as reflected in the core matrix (see section 6.9).

Our next step is to construct subject i's scores on a real variable j instead of on the latent variables q_1 and q_2 analoguously to the above procedure for subjects:
Subject i's score on real variable j under prototype condition r_1 is

$$\begin{aligned} v_{ij1} &= h_{jq_1} s_{iq_1 r_1} + h_{jq_2} s_{iq_2 r_1} \\ &= h_{j_1} s_{i_{11}} + h_{j_2} s_{i_{21}} \quad . \end{aligned}$$

Similarly, on real variable j under prototype condition r_2 we have

$$v_{ij2} = h_{j1} s_{i12} + h_{j2} s_{i22},$$

where the weights h_{j1} and h_{j2} indicate to what extent the latent variables determine the real variable j.

Finally we combine the prototype conditions. Subject i's score on test j under condition k may be written as

$$z_{ijk} = e_{k1} v_{ij1} + e_{k2} v_{ij2}$$

where the weights e_{k1} and e_{k2} indicate to what extent each prototype condition determines the real condition k.

Assembling the results from the three steps we get:

$$z_{ijk} = \sum_{r=1}^{2} e_{kr} v_{ijr} = \sum_{r=1}^{2} e_{kr} \{\sum_{q=1}^{2} h_{jq} s_{iqr}\}$$

which can be compactly written as

$$z_{ijk} = \sum_{r=1}^{2} e_{kr} \underbrace{\{ \underbrace{\sum_{q=1}^{2} h_{jq} [\underbrace{\sum_{p=1}^{2} g_{ip} c_{pqr}}_{\text{linear combination of subjects } p_1 \text{ and } p_2}]}_{\text{linear combination of variables } q_1 \text{ and } q_2} \}}_{\text{linear combination of conditions } r_1 \text{ and } r_2} \quad ,$$

or

$$z_{ijk} = \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} g_{ip} h_{jq} e_{kr} c_{pqr}, \text{ with } s=t=u=2$$

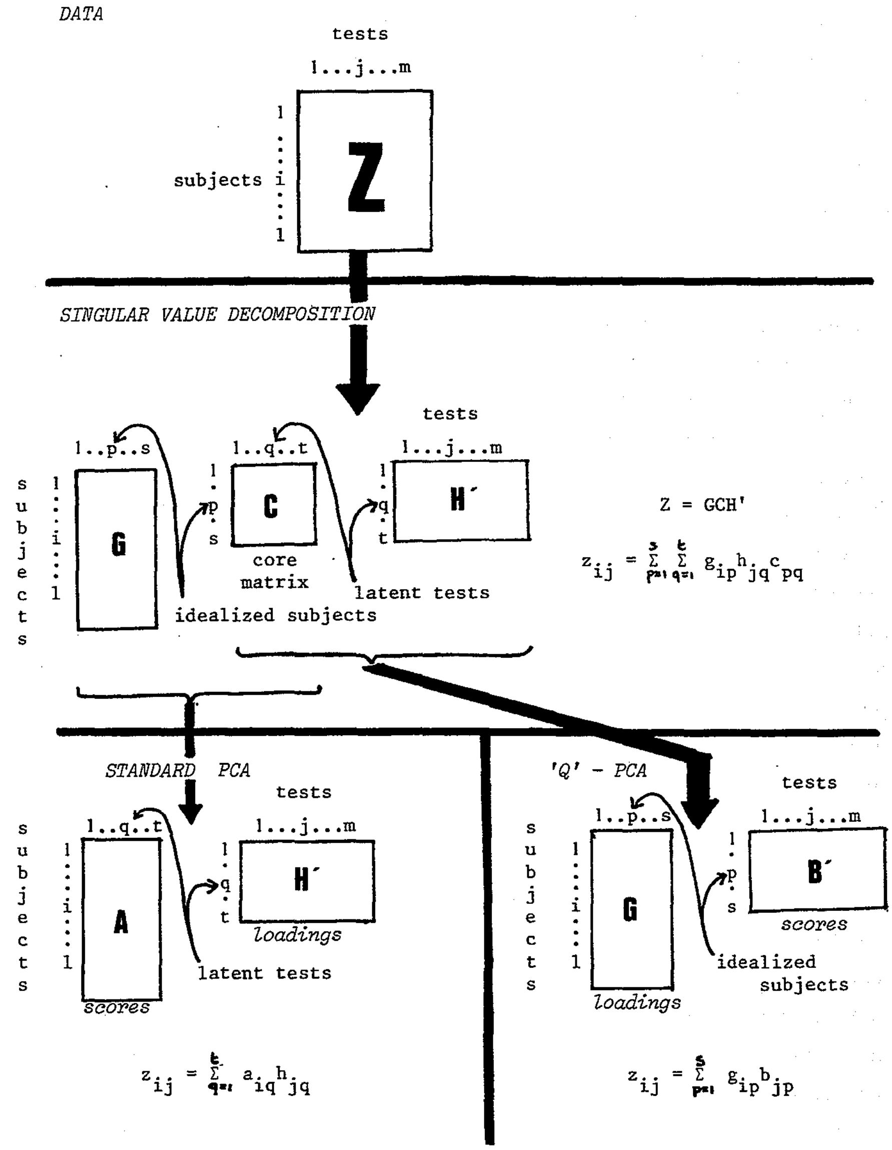


Fig. 2.1 *Singular value decomposition and standard principal component analysis*

as it is usually written. Note that the order in which the linear combinations are taken is immaterial.

As can be seen in section 2.3, this is the definition of the three-mode principal component model. In Bloxom (forthcoming) the three-mode model in its nested form is described as well, but there the model is developed as an example of a third-order factor analysis model, in which the s are the second order, and the v the third order factors.

Methodology: extending singular value decomposition. From a methodological point of view three-mode principal component analysis is a generalization of standard principal component analysis, or rather, of *singular value decomposition*. Fig. 2.1 schematically shows the relationship between standard principal component analysis and singular value decomposition. In essence, singular value decomposition is a simultaneous analysis of both the individuals and the variables, in which the interactions between the components of the variables and the subjects is represented by the core matrix C. In Fig. 2.1 the core matrix is diagonal with s diagonal elements $c_{pp}(p=1,...,s)$. These c_{pp} are equal to the square roots of the eigenvalues associated with the p-th components of both the variables and the subjects. When G and C are combined to form A, as shown in Fig. 2.1, we have the standard principal component solution, and when H and C are combined we have what could be called in Cattell's (1966a) terms 'Q'-principal component analysis.

Fig. 2.2 shows the decomposition of a three-mode matrix according to the three-mode principal component model. Comparison of Fig. 2.1 and Fig. 2.2 shows the analogy between the singular value decomposition and three-mode principal component analysis. The core matrix now has three modes, and the relationships between the singular values or elements of the core matrix and the eigenvalues of the various modes are less simple than in the two-mode case (see section 2.8, and section 6.9).

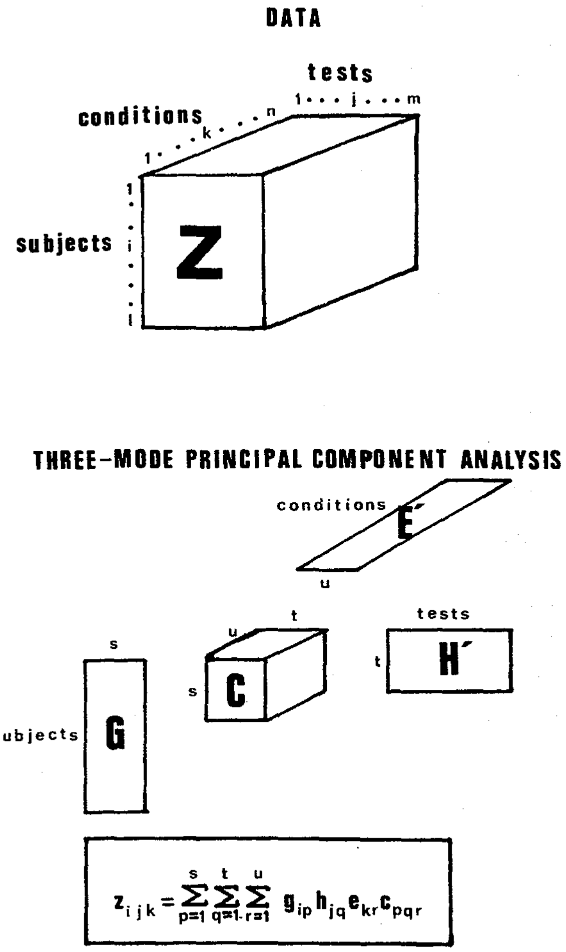


Fig. 2.2 *Three-mode principal component analysis*

2.3 FORMAL DESCRIPTIONS

In this section we present the Tucker3 and Tucker2 models; which form the basic working models for the rest of this book. The description here is rather superficial; a detailed treatment can be found in Chapter 4. Here we attempt to present just enough detail for understanding the main principles involved.

Tucker3 model. The general three-mode principal component model (or Tucker3 model) can be formulated as the factorization of the three-mode data matrix $Z = \{z_{ijk}\}$, such that

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

for $i=1,..,\ell$; $j=1,..,m$; $k=1,..,n$. The coefficients g_{ip}, h_{jq}, and e_{kr} are the entries of the component matrices G (ℓ×s), H (m×t), and E (n×u); ℓ,m,n are the number of elements, and s,t,u are the number of components of the first, second and third mode respectively. We will *always* assume that G, H, and E are columnwise *orthonormal* real matrices with the number of rows larger than or equal to the number of columns. The c_{pqr} are the elements of the three-mode core matrix C (s×t×u).

In practice the three-mode data matrix is not decomposed into all its components, as one is usually only interested in the first few. Therefore, one seeks an approximate decomposition $\hat{Z}$ that is minimal according to a least squares loss function, i.e one solves for a $\hat{Z}$ such that

$$\sum_{i=1}^{\ell} \sum_{j=1}^{m} \sum_{k=1}^{n} (z_{ijk} - \hat{z}_{ijk})^2$$

with

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

attains a minimum. The algorithm to solve this minimization problem is implemented in the program TUCKALS3 (Kroonenberg, 1981a). Details about the existence and uniqueness of a minimum, the algorithm itself, and its implementation can be found in Chapter 4.

Tucker2 model. An important alternative version of the Tucker3 model can be obtained by equating the component matrix E with the identity matrix. We will refer to this model as the Tucker2 model; it has also been called the generalized subjective metrics model (Sands & Young, 1980). It can be written as

$$z_{ijk} = \sum_{p=1}^{s} \sum_{q=1}^{t} g_{ip} h_{jq} \tilde{c}_{pqk},$$

or in matrix notation

$$Z_k = G\tilde{C}_k H' \qquad (k=1,..,n)$$

where $\underset{\sim}{Z}_k$ ($\ell \times m$) is the k-th *frontal plane* or slice of the data matrix, and $\tilde{C}_k$ (s×t) the *extended core matrix*, respectively. The core matrix is called extended because the dimension of the third mode is equal to the number of conditions in the third mode, rather than to the number of components as is the case in the Tucker3 model. The Tucker2 model only specifies principal components for the ℓ subjects and m variables but not for the n conditions. The relationships between the components of the subjects and the variables can be investigated for all conditions together as well as for each condition separately.

The least squares loss function for the Tucker2 model has the form

$$\sum_{i=1}^{\ell} \sum_{j=1}^{m} \sum_{k=1}^{n} (z_{ijk} - \hat{z}_{ijk})^2$$

with

$$z_{ijk} = \sum_{p=1}^{s} \sum_{q=1}^{t} g_{ip} h_{jq} \tilde{c}_{pqk},$$

and the algorithm to solve this minimization problem is implemented in the program TUCKALS2 (Kroonenberg, 1981c).

One important advantage of the methods discussed in this paper over the standard procedures outlined by Tucker (1966a, p.297ff.) is that the estimates of the parameters are least squares, rather than estimates with ill-defined properties. Another advantage of the definition of loss functions is that it becomes possible to look at residuals (see sections 2.4 and 2.9). A third advantage is that there exists a direct relationship between the eigenvalues of the configurations and the size of the elements in the core matrix (see sections 2.1 and 2.2; for an empirical demonstration see section 2.8). On the other hand, an advantage of Tucker's methods (which are described in section 4.2) is that they are cheap, and that their solutions are nested (see section 4.5 for a discussion of nesting).

2.4 INTERPRETATIONAL AIDS

Various kinds of auxiliary information can be useful for the interpretation of results from a three-mode principal component analysis. Some of the most important ones will be presented here, i.e. joint plots, component scores, use of residuals, scaling of input data, and rotations; some details will be taken up where needed in the example, and detailed discussions can be found in later chapters.

Joint plots. After the components have been computed, the core matrix will provide the information about the interactions between these components. For instance, it is very instructive to investigate the component loadings of the subjects jointly with the component loadings of the variables, by projecting them together into one space, as it then becomes possible to specify what they have in common. The plot of the common space is called a *joint plot.*

Such a joint plot of every pair of component matrices for each of the components of the third mode, say E, in the TUCKALS3 case, and for the average core plane in the TUCKALS2 case, is constructed in such a way that g_p $(p=1,...,s)$ and h_q $(q=1,...,t)$ - i.e. the columns of G and H respectively - are close to each other. Closeness is measured as the sum of all $s \times t$ squared distances $d^2(g_p,h_q)$ over all p and q. In section 6.10 we will discuss in some detail the construction of the joint plots.

Component scores. In some applications it is useful to inspect the scores of all combinations of the elements 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 (j) components can be used to inspect the development of an individual's score on the latent variable over time. The component scores also express the closeness of the elements from different modes in the corresponding joint plot. In the example presented in Chapter 8 these component scores in fact turn out to be the most successful summary of the relationships involved. In section 2.10 we also give an example of their usefulness, and they will be described in detail in section 6.10.

Residuals. In Chapter 4 it is shown that both for the Tucker3 and the Tucker2 model the following is true:

$$\sum_{i=1}^{\ell}\sum_{j=1}^{m}\sum_{k=1}^{n} z_{ijk}^2 = \sum_{i=1}^{\ell}\sum_{j=1}^{m}\sum_{k=1}^{n} \hat{z}_{ijk}^2 + \sum_{i=1}^{\ell}\sum_{j=1}^{m}\sum_{k=1}^{n} (z_{ijk} - \hat{z}_{ijk})^2 .$$

where the $\hat{z}_{ijk}$ are the data 'reconstructed' from the estimated parameters. This is, of course, a standard result in least-squares analyses. Less numerically this partitioning of the total sum of squares may be written as

$$\text{SS(Data)} = \text{SS(Fit)} + \text{SS(Residual)}.$$

In addition, it is shown that for each element f of a mode

$$\text{SS(Data}_f) = \text{SS(Fit}_f) + \text{SS(Residual}_f).$$

By comparing the fitted sum of squares and the residual sum of squares for the f-th element one can gauge the correspondence of the f-th element's configuration with the overall configuration. Large residual sums of squares indicate that a particular element does not fit very well into the structure defined by the other elements.

The size of the SS(Residual$_f$) of an element f of a mode generally depends on its SS(Total$_f$). Therefore, one should in general look at the *relative residual sum of squares* (or *relative residual*, for short), (= SS(Residual$_f$)/SS(Total$_f$)) when assessing the role of a particular element in the final solution. Similarly, one could look at the *relative residual* (= SS(Res$_f$)/SS(Total$_f$)). These two quantities convey essentially the same information.

The SS(Res) and the SS(Fit) as well as their relationships can be shown directly in a so-called *sums-of-squares plot,* which is explained and illustrated in Chapter 7. Section 2.9 (Fig. 2.5) also contains an example.

Scaling of input data. In standard principal component analysis the input data are often transformed into standard scores without much thought about the consequences. In other words, correlation matrices are generally analysed with principal component analysis, rather than cross-product matrices or covariance matrices. In three-mode analysis the question of scaling the input data must be approached with more care, as there are many ways to standardize or centre the data.

Two general rules can be formulated with regard to the scaling of input data:

- those means should be eliminated (i.e. set equal to zero), which cannot be interpreted, or which are incomparable within a mode;
- those variances should be eliminated (i.e. set equal to one) which are based on arbitrary units of measurement, or which are incomparable within a mode.

These rules do not lend themselves to automatic application. For each and every data set it has to be assessed which kind of scaling is most appropriate.

Very common procedures are (see section 6.5):

- centring and/or standardizing the variables over all subject-condition combinations *(j-centring)*, so that the grand mean of a variable over all subjects and conditions is zero, and/or its total variance over all subjects and conditions is one;
- centring and/or standardizing the variables over all subjects for each condition separately *(jk-centring)*;
- *double-centring*, i.e. centring per condition over both variables and subjects *(jk, ik-centring)*.

As before, subjects, variables, and conditions here indicate first, second, and third mode elements, respectively.

The decision which centring or standardization is appropriate with any particular data set depends on the researcher's assessment of the origin of the variability of the data, in other words, which means and variances can be meaningfully interpreted. In Chapter 6 a further discussion on this topic can be found.

2.5 PARTY SIMILARITY STUDY: DESIGN AND DATA

To familiarize the reader with some practical aspects of three-mode analysis, and to illustrate the main points of the previous sections we will analyse some out-of-date data on similarities between Dutch political parties collected by De Gruijter (1967).

De Gruijter employed 82 members of political student organizations at the University of Leiden, of whom three were not included in the present analysis. On the basis of their preference for a particular party the students were divided into six *preference groups*, viz. into a PSP, PvdA, KVP, ARP, VVD, and a CHU group. Table 2.1 based on Wolters (1975), contains a short characterization of the parties involved in this study.

The ten parties which were then in Parliament (1966) - CPN, PSP, PvdA, KVP, ARP, VVD, CHU, SGP, GPV, Boerenpartij (BP) - were used as stimuli. De Gruijter confronted the students with all possible triads of parties, and asked them to indicate for each triad which two parties were most alike, and which two were least alike. For each preference group he computed the number of times (summed over all students in that group - n_g) that in all triads with stimulus parties i and j, the similarity between i and j was considered to be greater than that between i (the *standard*) and a third stimulus. As each party was compared with all combinations of the other parties, the sums for the standards over all parties are equal to $n_g x\binom{9}{2}$. Thus, the data have the form of 6 matrices (one for each preference group) of 10 standards by 10 *compared parties*.

Unlike De Gruijter we have divided the data of each preference group by the number of students in that group (cf. Table 2.1), to eliminate the uninteresting differences between the groups due to their different sizes. In addition, the main diagonal elements of each matrix, which were left blank in De Gruijter's analysis, were set to 9 indicating that a party is more similar to itself than to any other party. Note that the data matrices can be, and are, asymmetric, as for a party there is no necessity to be as often considered alike to another party when compared with a standard as it is considered to be alike to that same party when the party itself is the standard. Note also that all row sums are now $\binom{10}{2} = 45$. The data matrices were double-centred (see Chapter 6) before the analysis proper, as is customarily done with similarities. Table 2.2 gives, as an example, the data (adjusted for group size and rounded to whole numbers) for the KVP and PvdA groups.

Table 2.1 *Description of Dutch parliamentary parties in 1966 and the numbers of first preferences for them among De Gruijter's student respondents*

	Name of party, initials	Description of party	number of preference
1	Communistische Partij Nederland (CPN)	communists	1
2	Pacifistisch-Socialistische Partij (PSP)	pacifists, radical left-wing socialists	8
3	Partij van de Arbeid (PvdA)	labor, social democrats	15
4	Katholieke Volks Partij (KVP)	roman catholics, Christian-Democratic party	11
5	Anti-Revolutionaire Partij (ARP)	protestant Christian-Democratic party; adherents are mainly members of 'gereformeerde' churches	10
6	Vereniging voor Vrijheid en Demokratie (VVD)	liberals, more conservative than British or German liberals	9
7	Christelijke Historische Unie (CHU)	protestant Christian-Democratic party; adherents are mainly members of 'hervormde' churches	9
8	Staatkundig Gereformeerde Partij (SGP)	protestant isolationist puritan calvinist party; adherents are mainly members of 'oud-gereformeerde' churches	0
9	Gereformeerd Politiek Verbond (GVP)	protestant nationalistic puritan calvinist party; adherents are mainly members of 'vrijgemaakt-gereformeerde' churches	1
10	Boeren Partij (BP)	poujadist-type of protest party	1

Source: Wolters (1975)

Table 2.2 *Party similarity study: Data of PvdA and KVP groups*

PvdA group
compared parties

standards	CPN	PSP	PvdA	KVP	ARP	VVD	CHU	SGP	GPV	BP
CPN	9	7	7	4	4	2	4	2	2	3
PSP	8	9	7	4	4	3	4	2	2	2
PvdA	6	7	9	6	6	3	$\boxed{5}$	2	1	1
KVP	1	2	6	9	6	6	6	2	3	3
ARP	1	2	5	5	9	$\boxed{4}$	7	4	$\boxed{5}$	2
VVD	0	2	3	6	5	9	7	4	4	$\boxed{5}$
CHU	0	1	4	5	7	6	9	$\boxed{4}$	5	3
SGP	1	2	3	3	$\boxed{6}$	4	5	9	8	6
GPV	1	2	2	3	$\boxed{6}$	4	6	8	9	6
BP	1	2	2	3	4	$\boxed{6}$	$\boxed{5}$	6	6	9

KVP group
compared parties

standards	CPN	PSP	PvdA	KVP	ARP	VVD	CHU	SGP	GPV	BP
CPN	9	8	7	4	3	2	3	3	2	$\boxed{5}$
PSP	7	9	6	3	3	2	4	4	3	$\boxed{5}$
PvdA	5	7	9	7	5	2	$\boxed{5}$	2	2	2
KVP	1	3	6	9	7	$\boxed{4}$	7	3	3	1
ARP	1	1	5	6	9	$\boxed{3}$	7	$\boxed{5}$	$\boxed{5}$	2
VVD	0	1	3	6	5	9	7	3	4	$\boxed{5}$
CHU	1	2	3	6	8	5	9	5	5	2
SGP	1	3	2	3	$\boxed{6}$	3	6	9	8	$\boxed{4}$
GPV	1	3	3	4	$\boxed{6}$	2	6	8	9	$\boxed{4}$
BP	4	4	2	3	3	$\boxed{5}$	4	5	6	9

$\boxed{i}$ similarity out of line with the main pattern

2.6 ANALYSES AND FIT

Analyses. The main analysis reported here, is a TUCKALS3 (T3) analysis with three components each for the first and the second mode (standards and compared parties), and with two components for the third mode (preference groups); this solution will be called the *3x3x2-solution*. It will sometimes be compared with another T3 analysis with two components for each of the modes, or the *2x2x2-solution*, and also with a TUCKALS2 (T2) analysis with two components for the first two modes, or the *2x2-solution*. As mentioned in section 2.3 no components are computed for the third mode in this model.

Fit. From Table 2.3 it can be seen that with three components for the party modes the variability in the data accounted for is 92% of the total sum of squares. Even the two component solutions are already satisfactory. The 'approximate fit' from the initial configuration for each of the modes (which are derived from the standard Tucker (1966a) Method I solution - see section 4.2) are upper bounds for the SS(Fit) of the simultaneous solution. Obviously the smallest of the three is the least upper bound, in this case the one based on the second mode (.94) - see also section 4.5. The initial configurations are used as starting points for the main TUCKALS algorithms. The improvement in fit indicates how much the iterative process improves the simultaneous solution over the starting solution. In this case this improvement is negligeable, in other words, we could have settled for the Tucker method as far as fit is concerned, but the changes in the component matrices might have been substantial even with small improvement in fit.

Table 2.3 *Party similarity study: Characteristics of solutions*

		T3 *3x3x2*	T3 *2x2x2*	T2 *2x2*
Standardized total sum of squares - SS(Total)		1.00	1.00	1.00
Approximation of SS(Fit) derived from separate PCA	on mode 1	.94	.83	.83
	on mode 2	.94	.83	.83
	on mode 3	.97	.97	--
Fitted sum of squares from simultaneous estimation		.92	.82	.82
Residual sum of squares from simultaneous estimation		.08	.18	.18
Improvement in fit compared to initial configuration		.001	.004	.000

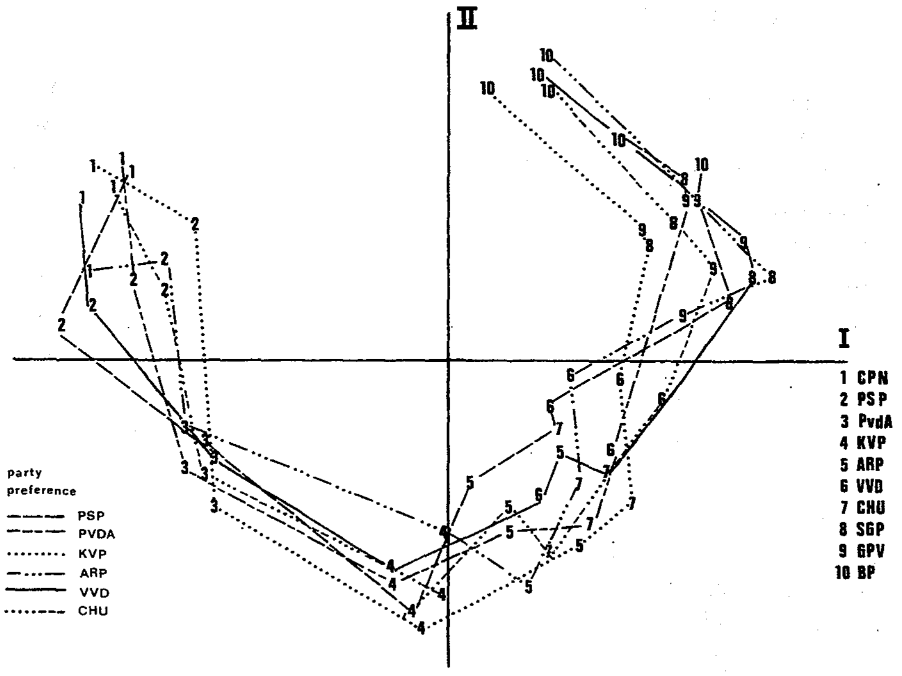


Fig. 2.3 *Party similarity study: Individual party spaces*

Table 2.4 *Party similarity study: Party spaces of mode 1 (standards)*

	3x3x2			2x2x2		2x2	
	1	2	3	1	2	1	2
CPN	.48	-.25	-.04	.48	-.25	.48	-.25
PSP	.48	-.17	.04	.48	-.18	.48	-.18
PvdA	.43	.18	.15	.43	.18	.43	.18
KVP	.01	.50	-.09	.01	.50	.01	.50
ARP	-.15	.30	.40	-.15	.30	-.15	.32
VVD	-.20	.22	-.62	-.20	.21	-.20	.20
CHU	-.22	.30	-.00	-.22	.29	-.22	.29
SGP	-.33	-.28	.34	-.33	-.27	-.33	-.27
GPV	-.33	-.33	.29	-.33	-.33	-.33	-.32
BP	-.17	-.46	-.47	-.17	-.46	-.17	-.47
component weight	.61	.21	.11	.61	.21	.61	.21

2.7 CONFIGURATIONS FOR THE THREE MODES

De Gruijter had to symmetricize his matrices due to the inability of earlier multidimensional scaling programs to handle asymmetric data. He analysed each preference group separately, rather than simultaneously as was done here. His results are displayed in Figure 2.3. The advantage of the present approach is that one space can be found for all groups together, and the fit of this common configuration for each group can be assessed. De Gruijter only extracted two dimensions, and concluded that a 'horseshoe' could be found for each preference group separately. It came as no surprise that in the present analysis the first two components of the common space exhibit a horseshoe as

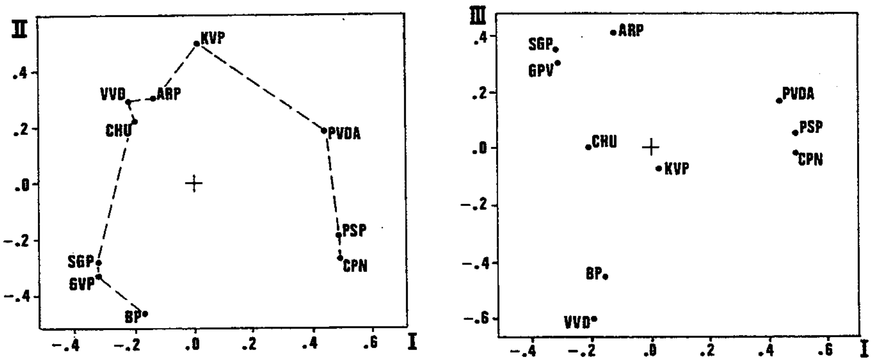


Fig. 2.4 *Party similarity study: Party space for standards* (3x3x2-solution)

well (see Figure 2.4). Table 2.4 shows that the first two components of all three solutions are virtually identical. As the differences between the fit of the models and the fit estimated from the separate principal component analyses on mode 1 is in all cases very small, the solutions should be similar.

So far we have implicitly assumed that we could look exclusively at the first mode solution. Of course we get a solution for the second mode as well, but the solutions for the standards and the compared parties are hardly different. In other words, the asymmetry present in the data is very small; a conclusion also reached by De Gruijter using different means.

Horseshoes, often found in multidimensional scaling, always pose problems for interpretation. Often but not always both the projections of stimuli on the axes, and their positions along the horseshoe, are candidates for interpretation. Guttman (1954), Kruskal & Wish (1978, p. 88,89), Borg (1976), and Gifi (1981, p. 246ff.) discuss horseshoes and their interpretations. In the present example the interpretation of the positions of the parties along the horseshoe is very clear cut, viz. from left-wing (CPN) to right-wing (BP).

The party space is open to a more complex interpretation than just the horseshoe. More particularly, the first dimension also shows a left-right distinction, so that we are now faced with the problem that we have not a priori scaled the parties on this dimension, and that we do not know which of the GPV, SGP, and BP the students considered the most right-wing party. Such information would have made it possible to choose between an interpretation on the basis of the horseshoe or the axis.

The second axis separates the big and ideologically or politically flexible parties with governmental experience from the small and more dogmatic parties which have never borne governmental responsibility. Which of the three characteristics mentioned the students really used, or used more often, is not possible to determine without additional information.

Finally the third axis, which is rather difficult to interpret, indicates that BP and VVD are alike, and both are unlike the ARP. It is possible that we are here fitting ideosyncracies of the data, but on the other hand the similarities mentioned can be observed in almost all data matrices. For each of the preference groups (for examples see Table 2.2) we see primarily a central band of high similarities which is responsible for the horseshoe. Not fitting into this pattern are exactly the relations represented by the third axis.

A further point worth mentioning is that the bending back of the horseshoe to make the BP somewhat alike to the CPN (probably because of their similar extremist and unflexible approach to politics) can be seen in the data by the slight increase in similarities in the upper-righthand and lower-lefthand corners of the data matrices. This effect is strongest for the KVP preference group, which is a party in the middle of the political spectrum, and least so for the PSP preference group near the end of the horseshoe (for a corroboration of this point see De Leeuw & Heiser, 1980, p.516,517). The circumplex structure (see e.g. Guttman, 1954; Shepard, 1978) is clearly not complete.

Table 2.5 *Party similarity study: Space of preference groups*

party preference	3x3x2		2x2x2	
	1	2	1	2
PSP	.42	.68	.42	.52
PvdA	.42	.34	.42	.22
KVP	.40	-.59	.40	-.79
ARP	.40	-.19	.40	-.09
VVD	.42	-.10	.42	.21
CHU	.38	-.18	.38	-.12
component weight	.91	.01	.81	.007

Finally we want to point out that the method used to solve the three-mode model precludes the solutions from being *nested,* i.e. the first two components of the 3x3x2-solution are not equal to the components of the 3x2x2-solution. That the difference is very small in the present case is beside the point (see also section 4.5).

The *party preference space* (Table 2.5) is on the whole hardly interesting. As was to be expected from the similarity of the solutions from the separate analyses by De Gruijter, the loadings on the first component are virtually equal. The second component accounts for only 1% of the total variation, and it probably reflects some very specific interactions which we will try to unravel in section 2.10. The importance of these interactions is slight, but it should be remembered that only the first and second modes were centred, and not the third. This means that the first component of the third mode still reflects the average scoring level of the six groups. Technically speaking, the average frontal plane of the double-centred three-mode matrix is not zero, while the average lateral and horizontal planes are (see also section 6.5; table 6.2). In the case under consideration, the second component of the third mode seems to reflect relatively unimportant aspects of the data, but with other data the second component might contain valuable information about the differences between the elements of the third mode, even though it is far smaller than the first component.

2.8 CORE MATRICES

TUCKALS3 core matrix. (Table 2.6). In this subsection we will only discuss the interpretation of the first frontal core plane. The core matrix indicates how the various components of the three modes relate to one another. For instance, the element c_{111} (=19) of the T3 core matrix indicates the strength of the relation between the first components of the three modes, and c_{221} (=11) the strength of the relation between the second components of the first and second modes in combination with the first of the third mode. The interpretation of the elements of the core matrix is facilitated if one knows that the sum over all squared elements of the core matrix is equal to the SS(Fit). In other words, the c^2_{pqr}'s indicate how much the combination of the p-th component of the first mode, the q-th component of the second mode, and the r-th component of the third mode contributes to the overall fit of the model, or how much of the total variation is accounted for by this particular combination of components. Thus as Table 2.6 shows, 60% of the SS(Total) is accounted for by the combination of the first components of the three modes, another 21% by c^2_{221}, and 11% by c^2_{331}. Together with the negligible contributions of the other elements of the first frontal plane these contributions add up to 91%, which is equal to the standardized weight of the first component of the third mode, as it should be. The core matrix thus breaks the SS(Fit) up into small parts, through which the (possibly) complex relations between the components can be analysed. It is in this way that we can interpret the core matrix as the generalization of eigenvalues or of the singular values of the singular value decomposition (see also Section 2.2 and 6.9). It constitutes a further partitioning of the 'explained' variation as is indicated by the eigenvalues of the standard principal component analysis.

The present example is in a way too simple to make full use of the interpretational possibilities of the core matrix, as all off-diagonal elements are virtually zero. In later examples, especially Chapter 8, it will be demonstrated more fully how the core matrix can be used, and in Chapter 6 we will give a detailed discussion of possible interpretations of the core matrix.

Table 2.6 *Party similarity study: Core matrix (TUCKALS3)* (frontal planes)

3x3x2-solution		core plane			standardized contribution to SS(Fit) of c_{pqr}			standardized contribution to SS(Fit) of plane
		components compared parties						
plane 1 (C_1)		1	2	3	1	2	3	
components	1	19	-0.03	-0.01	.60	.00	.00	.92
	2	0.06	11	-0.2	.00	.21	.00	
standards	3	0.01	0.03	8	.00	.00	.11	
plane 2 (C_2)								
components	1	0.5	1.3	1.2	< .003			.01
	2	1.1	-0.6	-0.4				
standards	3	0.9	-0.1	-0.4				

Note: The core matrix can be split up in two other ways, i.e. into horizontal and into lateral planes. The sums of the squared standardized elements of the horizontal planes are equal to the component weights or standardized contributions to the SS(Fit) of the *standards*. The analoguous sums of the lateral planes are equal to the component weights of the *compared parties*.

TUCKALS2 extended core matrix. The extended core matrix can be interpreted in essentially the same way as the TUCKALS3 core matrix in terms of the amount of explained variation. Again the sum of the squared elements equals the fitted sum of squares, but now the sum of the squared elements of a frontal plane, $\tilde{C}_k$, equals the contribution of the k-th element (party preference group) to this fitted sum of squares.

In those cases where two modes are equal or the components define the same space, as in the present example, another interpretation of the core matrix is possible. Within the context of multidimensional scaling of individual differences (see section 3.2 and 3.3) the input similarity matrices typically satisfy these conditions, and within

Table 2.7 *Party similarity study: Extended core matrix* (*TUCKALS2*)

Preference		core plane		standardized contribution to SS(Fit) $\tilde{c}_{pqk}$		standardized contribution to SS(Fit) plane k =st.SS(Fit_k)	relative fit
		1	2	1	2		
PSP	1	8.1	-0.8	.11	.00		
	2	-0.8	4.4	.00	.03	.14	.82
PvdA	1	8.0	-0.4	.11	.00		
	2	-0.5	4.7	.00	.04	.14	.86
KVP	1	6.9	0.9	.08	.00		
	2	0.5	5.5	.00	.05	.13	.81
ARP	1	7.5	0.1	.09	.00		
	2	0.2	4.5	.00	.03	.13	.80
VVD	1	8.5	-0.1	.12	.00		
	2	0.1	4.1	.00	.03	.15	.84
CHU	1	7.5	0.5	.09	.00		
	2	0.5	3.9	.00	.03	.12	.78
	1	7.8	0.0				
	2	-0.0	4.5	.60	.12		.82
		average core plane		standardized contribution to SS(Fit) of components of 1st and 2nd mode =st.SS(Fit_p) ≅ st.SS(Fit_q)			overall relative fit = st.SS(Fit) = Σst.SS(Fit_k) = Σst.SS(Fit_p) = Σst.SS(Fit_q)

this field an interpretation has been developed in terms of correlations and direction cosines of the axes of the spaces common to two (generally the first and second) modes(see Tucker, 1972a, p.7, and Carroll & Wish, 1974, p.91).

Given that the component spaces are the same, as is the case here, it makes sense to speak about the correlation or the angle between the first and second component of the common space. This angle can be derived from the off-diagonal elements of the core plane, if we assume that the differences between $\tilde{c}_{pqk}$ and $\tilde{c}_{qpk}$ are merely due to random fluctuations. An off-diagonal element can be looked upon as a direction cosine or correlation between component p and component q, provided $\tilde{c}_{pqk}$ is scaled, by dividing it by $\tilde{c}^{\frac{1}{2}}_{ppk}$ and $\tilde{c}^{\frac{1}{2}}_{qqk}$.

The direction cosines indicate the angles under which the k-th condition (party preference group) 'sees' the axes or components of the common space. In the present example the largest deviation from orthogonality is found for the PSP-group ($\cos \alpha_1 = \frac{1}{2}(0.8392 + 0.7604)/ 8.1051^{\frac{1}{2}} \times 4.4303^{\frac{1}{2}} = .134$; $\alpha_1 = 82^{o}$). It seems safe to assume that the deviations from orthogonality are more or less chance fluctuations.

2.9 FIT OF PREFERENCE GROUPS AND STIMULI

In essence the analysis could stop with the above interpretations. All that the technique has to offer towards breaking down complex relationships into small intelligible pieces is contained in the analysis so far. However, it is good to have some auxiliary information available to assess if there are no irregularities in the data such as outliers, unduly influential points, points which are not sufficiently accounted for, etc. An attractive way to investigate such questions is to inspect the residual sums of squares in conjunction with the fitted sums of squares (see section 2.4 and Chapter 7). Whereas the core matrix informs us about the contributions of the components and their interrelationships, the sums of squares broken down by the elements of the modes inform us about the contributions of these elements to the solutions. As it is our aim to show general principles we will only consider the TUCKALS3 solutions here.

The smaller SS(Total)s for the ARP, CHU, and KVP groups in Table 2.8 show that they tended to give slightly less outspoken judgements than the VVD and PSP groups. Although very often elements with larger total sums of squares tend to be fitted better than those with smaller sums of squares, this effect is hardly present in these data, possible due to the small differences in the SS(Total)s. The SS(Fit)s or, better, the relative fits are high and have about the same value, indicating that the common solution is shared by all to the same extent. The sums of squares for the third mode thus tell us that no really deviant groups are present. In a later paragraph we will slightly qualify this statement.

Table 2.8 *Party similarity study: Standardized sums of squares broken down by preference groups*

3x3x2-solution

Party preference	SS(Total)	SS(Fit)		SS(Res)	
	standardized	standardized	relative fit	standardized	relative res.
PSP	.18	.16	.93	.01	.07
PvdA	.17	.16	.95	.01	.05
KVP	.16	.15	.93	.01	.07
ARP	.16	.14	.91	.01	.09
VVD	.18	.16	.92	.01	.08
CHU	.15	.14	.91	.01	.09
overall	1.00		.92		.08
	st.SS(Total)	= st.SS(Fit)	+	st.SS(Res)	

Note: st. = standardized

Far more interesting is the comparison between the sum of squares for the *stimuli* (here shown for the first mode) in the 3x3x2-solutions and the 2x2x2-solutions (see Table 2.9). The total sums of squares sums of squares show that parties in the political centre have few high similarities with other parties, as one would expect for a party in the middle, and thus their total sums of squares are small. The parties at both ends of the spectrum are relatively more similar (remember that the similarities are centred) and thus have larger total sums of squares (see also the data matrices in Table 2.2).

The SS(Fit)s for the 2x2x2-solution show that the relative fit of the VVD, ARP, and the BP leaves much to be desired. The sums-of-squares plot (Fig. 2.5) summarizes most of the information of Table 2.9, and makes it easy to spot especially the ill-fitting and the well-fitting points. Furthermore, Table 2.9 shows that not much is gained by adding a fourth axis to the party space. Not only is the overall fit of the 3x3x2-solution very good, but all parties also fit more or less equally well. In other words, the third component was all that was needed to accommodate the remaining anomalies.

Table 2.9 *Party similarity study: Standardized sum of squares broken down by 'standards'*

	3x3x2			2x2x2			
	SS (Total)	SS(Fit)		SS(Fit)			Improvement in relative fit due to third component
Party	standardized	standardized	relative fit	standardized	relative fit	residual/fit ratio	
CPN	.16	.16	.95	.15	.94	.06	.01
PSP	.15	.15	.95	.15	.95	.05	.00
PvdA	.13	.12	.94	.12	.92	.09	.02
KVP	.06	.05	.84	.05	.82	.22	.02
ARP	.06	.05	.88	.03	.59	.71	.29
VVD	.08	.08	.93	.03	.41	1.43	.52
CHU	.06	.05	.83	.05	.83	.21	.00
SGP	.10	.10	.95	.09	.85	.17	.10
GPV	.10	.10	.95	.08	.82	.22	.13
BP	.09	.09	.90	.06	.66	.53	.24
overall	1.00		.92		.82	.09	.10

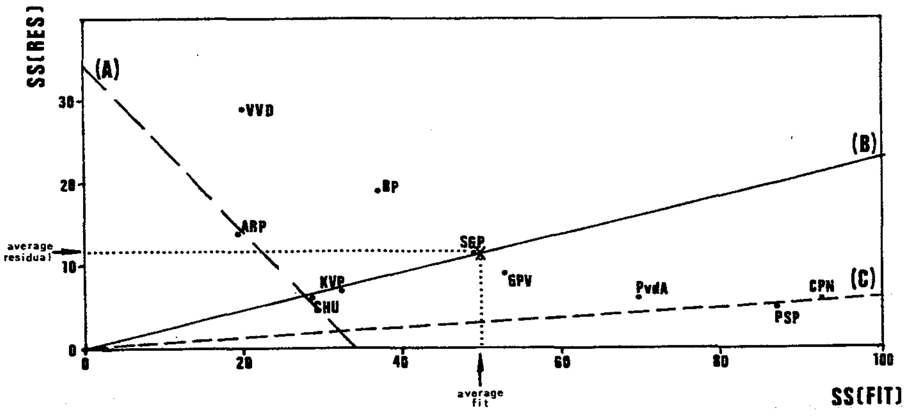


Fig. 2.5 *Sums-of-squares plot for standards (2x2x2-solution)*
(A): line at an angle of -45°; locus of points with equal total sums of squares;
(B): line connecting (0,0) and (average SS(Fit), average SS(Res)); separating the well-fitting points (below the line) from the ill-fitting points (above the line)
(C): line with equal Rel.SS(Fit), and equal residual/fit ratio

2.10 JOINT PLOTS AND COMPONENT SCORES

Figure 2.6 shows a *joint plot* (see sections 2.4 and section 6.10) of the first and second mode, based on the first component of the third mode. In this case, all it tells us is that the original data are virtually symmetrical, which is by now no longer a surprise. The only thing worth mentioning seems to be that the discrepancy is larger for the KVP and the PvdA, but whether this is really important, it is difficult to assess. For many other data sets, however, in which the first and second component are different types of elements, these joint plots are a major aid in interpretation as can, for instance, be seen in almost all the examples in later chapters.

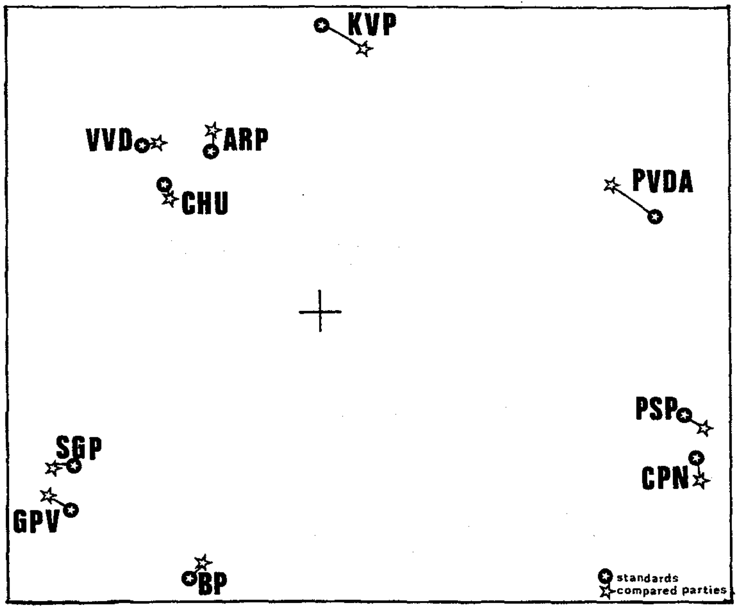
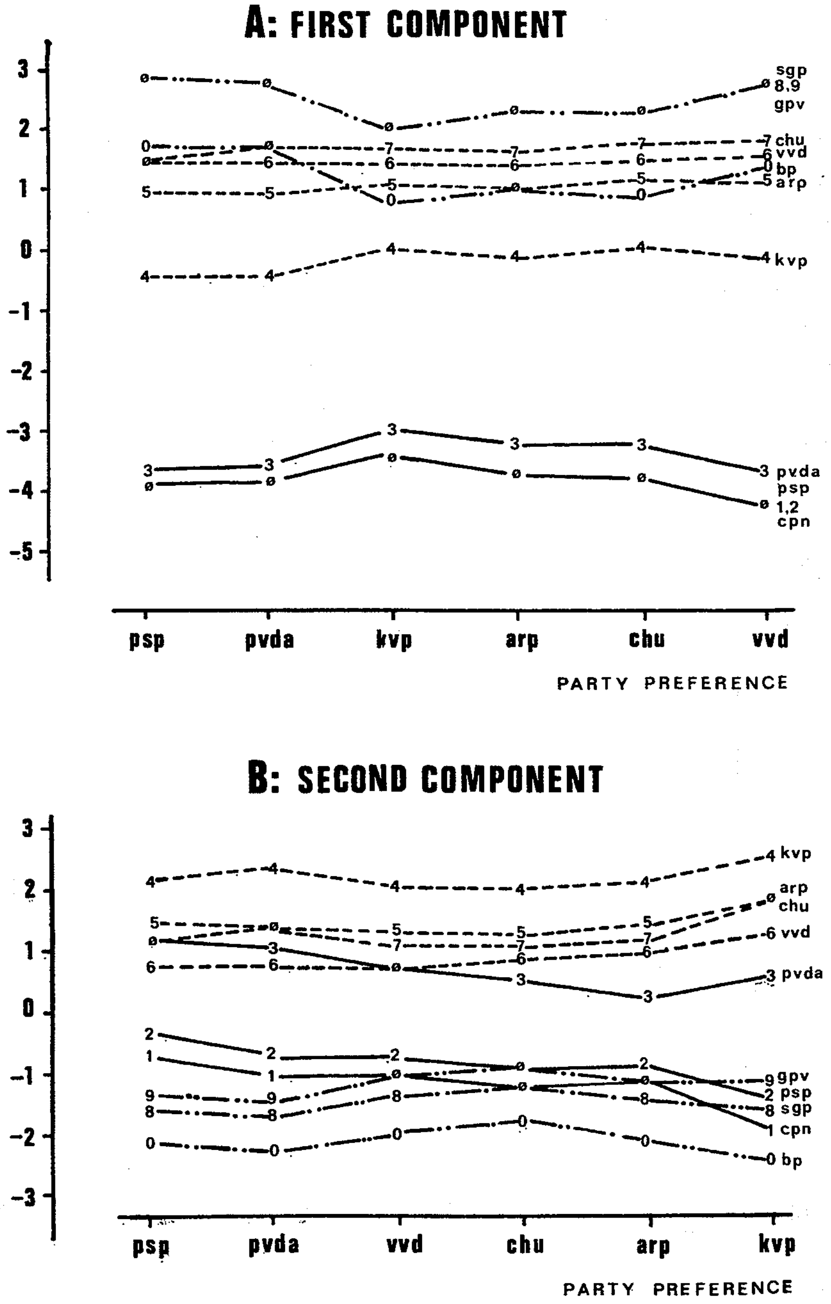


Fig. 2.6 *Party similarity study: Joint plot of standards and compared parties*

Component scores in three-mode analysis can sometimes serve as an intermediate level of condensation between the raw data and the model expressed in the component loadings and the core matrix. For the present data this is demonstrated in Figure 2.7A and B. The component



A: FIRST COMPONENT

B: SECOND COMPONENT

Fig. 2.7 *Party similarity study: Component scores on second mode components* scores for each standard-preference group combination on the first and second component of the party space (here defined by compared parties) are shown for the 2x2x2-solution. The different orders for the preference groups were chosen to achieve greater clarity of the figures.

In Figure 2.7A we see the expected similarity in judgement of the preference groups with respect to the relative position and numerical values for the standards. However, we also see the interaction of BP with preference group (PSP, PvdA: BP ≅ CHU, VVD; others: BP ≅ ARP) on the first component. Furthermore, PSP, PvdA, and VVD emphasize the left-right axis somewhat more than the Christian Democrats (KVP, ARP, CHU). This difference corresponds to the differences in the first component of the party preference space.

The important interaction between party preference and standards is the different treatment by the preference groups of the 'dogmatic, small' left-wing parties (PSP, CPN), and to a lesser extent of the PvdA, as is shown in Figure 2.7B. In particular, the left-wing preference groups (PSP, PvdA) make a clear distinction between the small left-wing (PSP, CPN) and the small conservative parties (GPV, SGP, BP) on the second component of the party space. Note also the profile similarity of the left-wing parties: the 'big, flexible' PvdA resembles the 'small,dogmatic' CPN and PSP. Especially these interactions create the second component of the space of the preference groups as is clear from Table 2.5: the interaction is large for the PSP and PvdA groups on the positive side, and large for the KVP on the negative side.

Reinspecting the core matrix of the TUCKALS3 solution (Table 2.6) we see that in the second plane (according to the second preference group component) a -be it slight- interaction is indicated between the left-right and the 'flexible,big' - 'dogmatic,small' dimensions, which demonstrates that this core plane contains information about the interactions. However, it is clear that unravelling and interpreting them is far from straightforward.

I

THEORY

SUMMARY

After presenting a survey of the organization of this book, its terminology, and notation in *Chapter 1*, and a survey of the subject matter of the book in *Chapter 2*, the main theory of three-mode principal component analysis is presented in the three chapters of *Part I*.

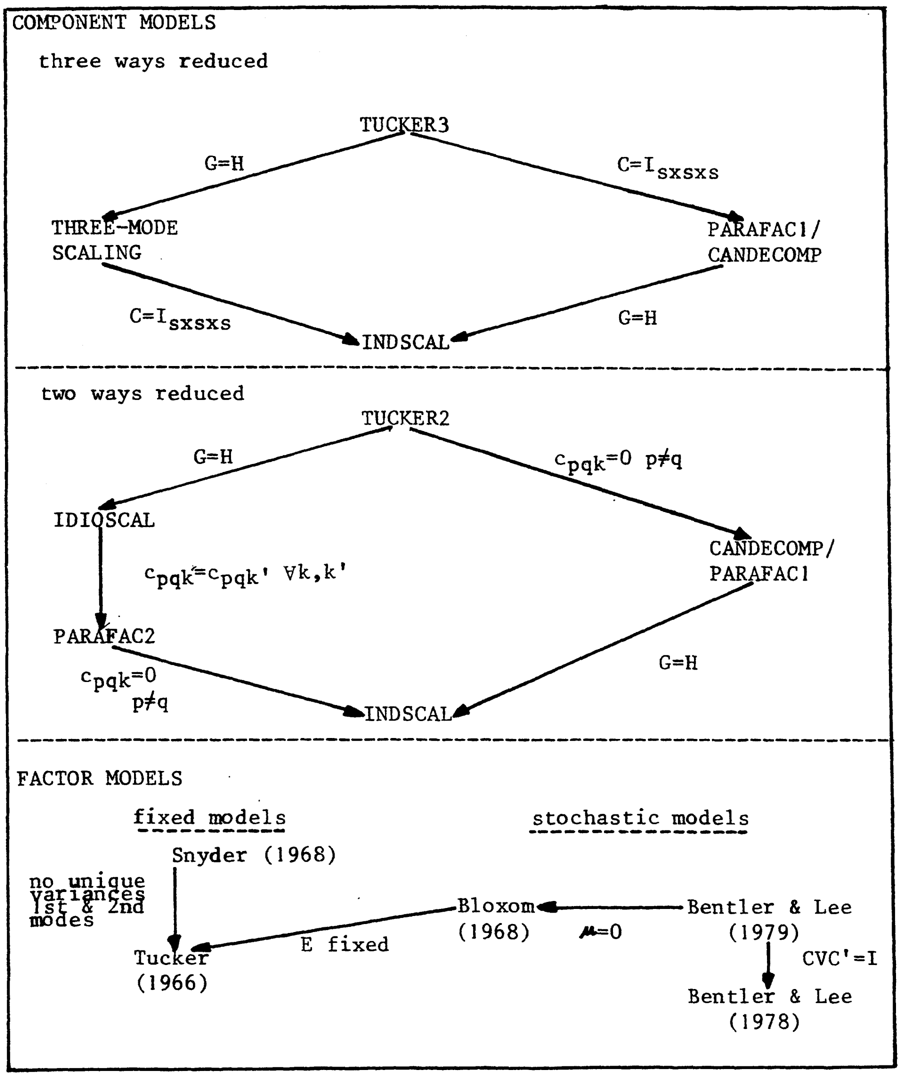
Chapter 3 deals with various models commonly used for the analysis of three-mode data. Two basic types may be distinguished: component models and factor analysis models. In the former all modes are considered fixed, whereas in the latter models the mode of observational units is considered stochastic. Within the class of component models two subclasses exist: models with three reduced modes (e.g. the Tucker3 model, three-mode scaling, PARAFAC1, and INDSCAL), and models with two reduced modes (e.g. the Tucker2 model, PARAFAC2, IDIOSCAL, CANDECOMP, and INDSCAL). These models are generally solved with alternating least squares procedures. The factor analysis models are generally solved within the context of the analysis of covariance structures. Chapter 3 ends with a number of extensions of the Tucker models, e.g. missing data facilities, optimal scaling procedures, facilities for external analysis, etc.

Chapter 4 presents the main theory connected with alternating least squares (ALS) solutions for fitting the Tucker models. Topics discussed include: existence of exact and approximate solutions, construction of ALS algorithms, convergence of the algorithms, as well as a number of other technical details. Three small data sets are used to assess the accuracy and correctness of the algorithms, and to investigate the robustness with increasing error in the data.

Finally, in *Chapter 5* transformation procedures are considered which aim to find 'simple' structures in core matrices. In particular algorithms for orthonormal and non-singular transformations are presented, illustrated, and evaluated.

MODELS

3


COMPONENT MODELS
three ways reduced
TUCKER3
G=H
C=I sxsxs
THREE-MODE SCALING
PARAFAC1/ CANDECOMP
C=I sxsxs
G=H
INDSCAL
two ways reduced
TUCKER2
G=H
c pqk=0 p≠q
IDIOSCAL
CANDECOMP/ PARAFAC1
c pqk=c pqk' ∀k,k'
PARAFAC2
c pqk=0 p≠q
G=H
INDSCAL
FACTOR MODELS
fixed models
stochastic models
Snyder (1968)
no unique variances 1st & 2nd modes
Bloxom (1968)
μ=0
Bentler & Lee (1979)
Tucker (1966)
E fixed
CVC'=I
Bentler & Lee (1978)

3.1. INTRODUCTION

In this chapter we will first give a survey of the plethora of models that have been proposed for the analysis of three-mode data (sections 3.2 through 3.6). We will restrict ourselves, however, to those models which explicitly or implicitly include a core matrix, and which primarily treat metric, unconditional data. Hereby we exclude such models or methods as simultaneous factor analysis (Jöreskog, 1971), longitudinal factor analysis (Corballis, 1973; Jöreskog, 1979; Jöreskog & Sörbom, 1977), point-of-view analysis (Tucker & Messick, 1963), three-mode point-of-view analysis (Tzeng & Landis, 1978), and three-way unfolding (DeSarbo, 1978; DeSarbo & Carroll, 1979, 1981).

In section 3.7 and following we will discuss a variety of extensions of the basic three-mode principal component models as well as a number of developments in related fields which have relevance for three-mode principal component analysis. We have grouped the extensions around three themes: 1. direct extensions of the basic models, viz. inclusion of missing data, extensions to other measurement levels, external analysis, and restrictions on the configurations (section 3.7); 2. three-mode causal modelling (section 3.8); 3. *n*-mode extensions (section 3.9).

The models considered in this chapter (see Fig. 3.1) are either scalar-product or Euclidean distance models. The latter will, however, only be treated in their scalar-product form, but this entails no loss of generality as the (generalized) Euclidean distance between two vectors can be defined as the (generalized) scalar-product between the difference vector and itself (see Carroll & Wish, 1974, p.183, 186; De Leeuw & Pruzansky, 1978, p.480).

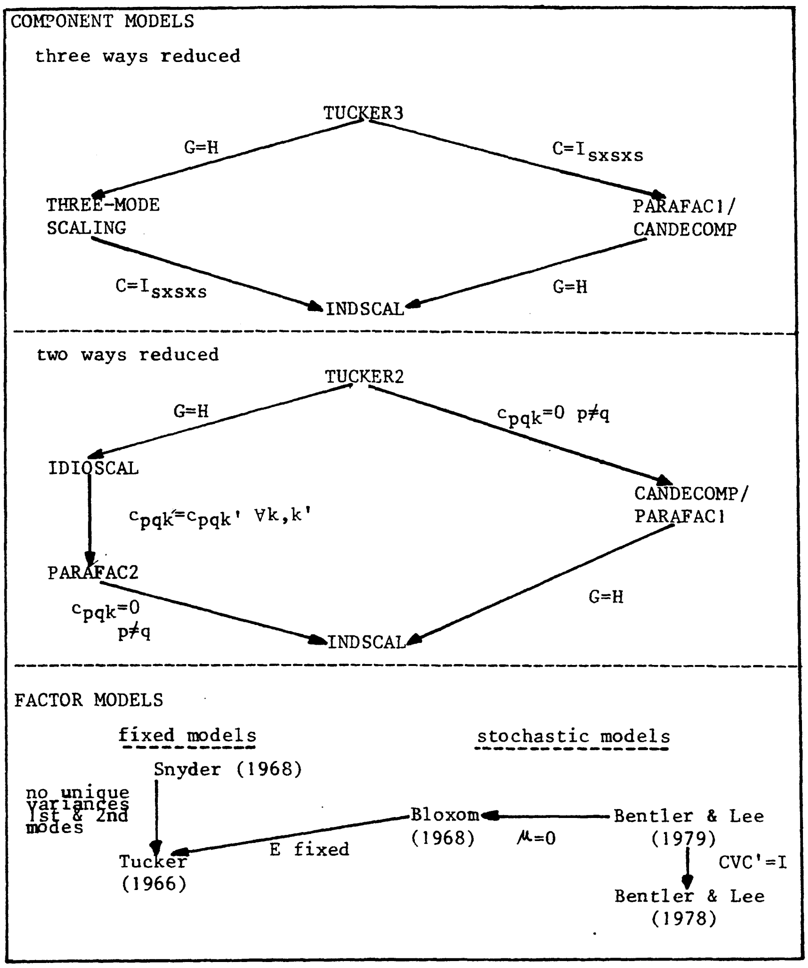


Fig. 3.1 *Model trees*

Two basic types of models are distinguished:

1. 'component' models or individual differences models, and
2. common factor (analysis) models or covariance structure models

The first type of models is considered determinate or fixed in all three modes, whereas the second type has one mode (usually individuals) which is stochastic. In fact, Lohmöller & Wold (1980) briefly describe a third type with two stochastic modes, but we will not discuss this model. The 'component' models are generally

Table 3.1 *Component or individual differences models*

model	number of modes diff.	red.	size core matrix	restrictions on core matrix	number of data points	number of parameters	original proposers
			three-way reduced models				
Tucker3 model	3	3	$s \times t \times u$		ℓmn	$stu+\ell s+mt+nu$	Tucker(1963,1964,1966a)
Three-mode scaling	3	2	$s \times s \times u$	f-symmetric	$\frac{1}{2}\ell(\ell+1)n$	$\frac{1}{2}s(s+1)u+2\ell s+nu$	Tucker(1972a)
PARAFAC1/CANDECOMP	3	3	$s \times s \times s$	3-way identity	ℓmn	$\ell s+ms+ns$	Harshman(1970), Carroll & Chang(1970)
INDSCAL	3	2	$s \times s \times s$	3-way identity	$\frac{1}{2}\ell(\ell+1)n$	$\ell s+ns$	Carroll & Chang(1970)
			two-way reduced models				
Tucker2 model	2	3	$s \times t \times n$		ℓmn	$stn+\ell s+mt$	Israelsson(1969), Tucker(1975)
IDIOSCAL	2	2	$s \times s \times n$	f-symmetric	$\frac{1}{2}\ell(\ell+1)n$	$\frac{1}{2}s(s+1)n+\ell s$	Carroll & Chang (1970, 1972)
PARAFAC2	2	2	$s \times s \times n$	f-symmetric dir.cos.eq.	$\frac{1}{2}\ell(\ell+1)n$	$\frac{1}{2}s(s-1)+ns+\ell s$	Harshman (1972a)
CANDECOMP/PARAFAC1	2	2	$s \times s \times n$	f-diagonal	ℓmn	$ns+\ell s+ms$	Carroll & Chang(1970), Harshman(1970)
INDSCAL	2	2	$s \times s \times n$	f-diagonal	$\frac{1}{2}\ell(\ell+1)n$	$ns+\ell s$	Carroll & Chang(1970)

f-symmetric = symmetric in all frontal planes; f-diagonal = diagonal in all frontal planes; dir.cos.eq.= direction cosines equal; red.= reduced; diff.= different.

more 'data-analytic' and exploratory, while the factor analysis models are more statistical and confirmatory, but the distinction is not absolute.

The term 'component' has been put between quotes so far, because the word is generally used for orthogonal vectors which result from an eigenvalue-eigenvector decomposition of a sums-of squares-and-cross-products matrix, covariance or correlation matrix. In this chapter we will use the term loosely to refer to the columns of the matrices G, H, and E (see section 1.5 for the general notation in this book). In the Tucker3 and Tucker2 models which form the core of this book, these matrices are columnwise orthogonal, even orthonormal, but for other models in this chapter this is not necessarily the case.

Maybe a word is in order about the use of the term 'model'. The word has many different meanings and connotations for different people. Here we will use it in a loose sense, mainly referring to a factorization of the data into a number of matrices, which describe in a parsimonious way relations between variables or sets of variables. It is not implied that there is always an underlying theory which prescribes the data generating process. Thus, in our usage of the term 'model' it is not only factor analysis but also principal component analysis which deals with models.

In Table 3.1 most of the models proposed within the class of component models are given. They are arranged in two hierarchies running from the more general to the more specific. In the first set all models have three reduced ways, while in the second set they have only two. The most specific models, CANDECOMP and INDSCAL, fit into both sequences as they can be given two different representations, as follows from the formulas given below. Not all conceivable models within the classifications have been proposed, and it seems unnecessary to include them just for the sake of completeness.

3.2 COMPONENT MODELS WITH THREE REDUCED MODES

Tucker3 model. The Tucker3 model is formulated as

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

In his exposition, Tucker (1966a, p. 286) first developed this model with full column rank component matrices G, H, and E. Solving the model, Tucker used for G, H, and E columnwise sections of orthonormal matrices. The (columnwise) orthonormality of G, H, and E can be assumed without loss of generality. Given that a solution for orthonormal G, H, and E has been found, the component matrices may be transformed via non-singular transformations without additional loss in fit of the model, provided the core matrix C is counterrotated (see Chapter 5 for details). The orthonormality restriction on the component matrices has the advantage of simplifying the algorithm to solve the models.

The approach described here, i.e. finding the optimal solution for the Tucker3 model using the orthonormality for identification of the equations, and dealing with the transformation freedom afterwards, is a common one in principal component analysis. For instance, De Leeuw & Pruzansky (1978) advocated this approach to solve the INDSCAL model. Unless specifically mentioned otherwise, we will assume that G, H, and E are columnwise sections of orthonormal matrices when we refer to the Tucker3 (or T3) model.

The component matrices here represent, as below, the *loadings* of the elements of the modes on their respective components. The core matrix gives the scores of each component of one mode on each component of a second mode for each component of the third one. In other words the core matrix tells uw how closely the various components of the different ways are related. A more detailed discussion of the ways to interpret core matrices is given in the section 6.9.

Three-mode scaling. The formulation of three-mode scaling

$$z_{ii'k} = \sum_{p=1}^{s} \sum_{p'=1}^{s} \sum_{r=1}^{u} g_{ip} g_{i'p'} e_{kr} c_{pp'r}$$

specifies that G and H of the general model should be equal. This equality forces the core matrix to be symmetric in its frontal planes, i.e. for each r $c_{pp'r} = c_{pp'r}$ for all p and p'. As there is only one component space G for the first and second mode, a further interpretation of the off-diagonal elements (i.e. $c_{pp'r}$, $p \neq p'$) of the frontal planes becomes possible in terms of direction cosines between the axes of the component space. More specifically, the direction cosine is $c_{pp'r}/(c_{ppr})^{\frac{1}{2}}(c_{p'p'r})^{\frac{1}{2}}$, making it possible to assess how the angles between the axes of the component space are 'viewed' by each component of the third mode. In other words, the frontal planes can be interpreted as the transformation, each component of the third mode applied to the component space G, a transformation consisting of a non-singular transformation (non-zero off-diagonal elements) and differential stretching or shrinking (by $c^{\frac{1}{2}}_{ppr}$) of the axes of G (see section 6.9 for further discussion of this point).

PARAllel FACtor analysis (PARAFAC1). The formulation of this model

$$z_{ijk} = \sum_{p=1}^{s} g_{ip} h_{jp} e_{kp}$$

shows that it can be obtained from the general Tucker3 model by deleting the core matrix from the model or, similarly, allowing for the three-way analogue of the identity matrix (s×s×s). In this form the model was formulated by Harshman (1970, 1976), while the same model (CANDECOMP) with a different rationale (see below) was proposed by Carroll & Chang (1970); see also Carroll & Wish (1974, p.90-92, and section 6.2) for an exposition of the two approaches. The interpretation of the present form is taken from the idea of parallel proportional profiles (Cattell, 1944, p.84; Harshman, 1970, esp. p.15-19). If we let z_{ijk} be the score of the i-th subject on the j-th variable in the k-th condition, then the contribution of the p-th component to a given score is the product of the loadings of the respective subject, variable and condition on this component. Put differently, g_{ip}, h_{jp}, e_{kp} indicate the proportional influence of that loading on the score z_{ijk}. It is interesting to note that PARAFAC1 and the (general) Tucker3 model are the only 'symmetric' models, in the sense that all ways are treated in the same way.

INDividual Differences SCALing (INDSCAL). When seen as a specialization of PARAFAC1 the model can be written as

$$z_{ii'k}=\sum_{p=1}^{s} g_{ip}g_{i'p}e_{kp}$$

In this form the model does not provide many additional interpretations. It is simply a specialization of the parallel profile analysis with the first two modes equal. The model will be discussed in an alternative formulation in the next section.

3.3 COMPONENT MODELS WITH TWO REDUCED MODES

Typical for component models with two reduced modes is that the core matrix for the third mode has the same dimension as the original data matrix. For each element of the third mode (subjects, conditions, occasions, etc.) the relations between the components of the other two modes are given, i.e. the third mode is not reduced in the s×t×n *extended core matrix*. The information for each element of the third mode is contained in its frontal plane $\tilde{C}_k = \{\tilde{c}_{pqk}\}$. Tucker (1972a) calls the $\tilde{C}_k$'s 'individual characteristic matrices', but they will be referred to here by a more neutral name as *(frontal) core planes*. The distinction between the models to follow lies mainly in the restrictions they impose on the extended core matrix.

Tucker2 model. As the formulation

$$z_{ijk}=\sum_{p=1}^{s}\sum_{q=1}^{t} g_{ip}h_{jq}\tilde{c}_{pqk}, \quad \text{or } Z_k = G\tilde{C}_kH'$$

shows, the difference between the Tucker2 model and the Tucker3 model is the treatment of the third mode. Especially in those cases in which it does not make sense to compute components, over the third mode, the model may be fruitfully used. Unless specifically mentioned otherwise, we will assume (without loss of generality) that the component matrices G and H are columnwise sections of orthonormal matrices, when we refer to the 'Tucker2 (or T2)' model.

Individual Differences In Orientation SCALing (IDIOSCAL). The main restriction on the IDIOSCAL model

$$z_{ii'k} = \sum_{p=1}^{s} \sum_{p'=1}^{s} g_{ip} g_{i'p'} \tilde{c}_{pp'k}, \text{ or } Z_k = G\tilde{C}_k G'$$

is that the component matrices for the first and second mode are equal. The model is thus particularly appropriate when the first two modes refer to the same variables, stimuli, etc. to classify the data. As a consequence, the frontal planes of the core matrix will also be symmetric. The interpretation of this model is that the private space of each individual k is an orthogonally rotated version of the common space G, in which the dimensions of its own coordinate system are weighted by the square root of the diagonal elements of the core plane, i.e. $\tilde{c}_{ppk}$, also called *saliences* (Carroll & Chang, 1970).

PARAllel FACtor analysis (PARAFAC2). Restricting the core matrix even more by demanding that the off-diagonal elements should be equal, we get

$$z_{ii'k} = \sum_{p=1}^{s} \sum_{p'=1}^{s} g_{ip} g_{i'p'} \tilde{c}_{pp'k}, \text{ or } Z_k = G\tilde{C}_k G'$$

with $\tilde{c}^*_{pp'k} = \tilde{c}^*_{pp'k'}$ for k,k' = 1,...,n, and $p \neq p'$ and $\tilde{c}^*_{pp'k} = \tilde{c}_{pp'k}/\tilde{c}^{\frac{1}{2}}_{ppk}\tilde{c}^{\frac{1}{2}}_{p'p'k}$. The interpretation of this model is that all elements of the third mode 'view' the common axes under the same angles. The direction cosines of those angles, $\tilde{c}^*_{pp'k}$, are estimates of the *subjective intercorrelations* as Carroll & Wish (1974, p.96) call them. Thus the possibly oblique orientation is fixed over all elements of the third mode, but the axes may be weighted differently. Again the solutions for each third mode element are parallel, in the sense that the private spaces are all proportional to the common space but with different weights. Harshman (1972b) has made some claims about the uniqueness of the solution of this model, which were questioned by Carroll & Wish (1974, p.96) and which according to Carroll & Arabie (1980) were not settled at the time they finished writing their general review of multidimensional scaling models.

CANonical DECOMposition (CANDECOMP). Within the framework of extended core matrices CANDECOMP can be written as

$$z_{ijk}=\sum_{p=1}^{s} g_{ip}h_{jp}\tilde{c}_{ppk}, \text{ or } G\tilde{C}_kH',$$

with $\tilde{C}_k$ diagonal (k=1,...,n). It is mathematically equivalent to the PARAFAC1 model of the previous section (see also section 6.2). The interpretation of the CANDECOMP model is that the number of components for the first two modes should be equal and that the core matrix is restricted to be diagonal in all its frontal planes. For all third mode elements a one-to-one relation is postulated between the components of the first two modes, and they only differ in the weights or saliences they attach to each of the common dimensions. It has been shown (Harshman, 1972b; Kruskal, 1976,1977; De Leeuw & Pruzansky, 1978) that under rather weak conditions, a normalization such that the centroid of the stimulus spaces is at the origin, and the sum of the squared projections on each dimension in that space is one, the solution of CANDECOMP is unique except for permutations followed by a diagonal transformation. The importance of this uniqueness theorem is that if the model fits well the good fit only refers to the solution found, and not to infinitely many of them, which could be found by non-singular or orthogonal transformations.

INDividual Differences SCALing (INDSCAL). Although not generally written this way, our formulation of the model

$$z_{ii'k}=\sum_{p=1}^{s} g_{ip}g_{i'p}\tilde{c}_{ppk}, \text{ or } Z_k=G\tilde{C}_kG'$$

with $\tilde{C}_k$ diagonal (k=1,...,n), exhibits most clearly the essential features of the scalar-product form of the model. As in CANDECOMP the frontal planes are diagonal, and the interpretation of the model is that the third mode elements share a common (stimulus) space, the axes of which they may weight differently, viz. by a factor $\tilde{c}^{\frac{1}{2}}_{ppk}$. The diagonality of the core planes implies that the private spaces of the third mode elements are not rotated versions of the common space, but that they are all orientated in the same manner. The uniqueness theorem mentioned above holds a fortiori for the INDSCAL model.

3.4 GENERALITY OF THE TUCKER3 MODEL[$]

In this section we will discuss how general the Tucker3 model is compared to models like PARAFAC1/CANDECOMP and the Tucker2 model.

The standard matrix formulation of the Tucker3 model used in this book is

$$Z = GC(H'\otimes E').$$

Another way of presenting the model would be to specify it for each frontal plane Z_k:

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

Suppose that $C_1,\ldots,C_u$ can be diagonalized simultaneously, i.e. there exist non-singular matrices A and B, such that

$$D_r = AC_rB' \qquad (r=1,\ldots,u)$$

are diagonal at the same time (see Chapter 5 for a procedure to find such an A and B). If the number of components of the first and second mode, s and t respectively, are unequal then D_r is considered to be diagonal if all $d^r_{k\ell}$ are zero unless $k=\ell$. If s is not equal to t and D_r is diagonal in the sense described above, then one may assume without loss of generality that D_r is an axa matrix with a= min(s,t), as the (max(s,t)-a) components of one of the modes do not contribute to the fit of the model. Therefore, we will assume from now on that s=t in this section.

[$] This section is based on an informal note by Jan de Leeuw (1982)

We may now rewrite the model as

$$Z_k = G\{\sum_{r=1}^{u} e_{kr}C_r\}H' = GA^{-1}\{\sum_{r=1}^{u} e_{kr}AC_rB'\}(HB^{-1})' =$$

$$= U\{\sum_{r=1}^{u} e_{kr}D_r\}V' = UW_kV'$$

with $U = GA^{-1}$, $V = HB^{-1}$, and $W = \sum_{r=1}^{u} e_{kr}D_r (k=1,\ldots,n)$.

As all W_k are diagonal we may rewrite the model as

$$z_{ijk} = \sum_{p=1}^{s} u_{ip}v_{jp}w_{ppk} = \sum_{p=1}^{s} u_{ip}v_{jp}\tilde{w}_{kp}$$

with $\tilde{W} = \{\tilde{w}_{kp}\}$ is an (n×s) matrix whose rows are the diagonals of the W_k. This equation has the form of the PARAFAC1/CANDECOMP model (see section 3.2 and 3.3). In other words, if the core matrix from a Tucker3 model can be diagonalized, then the model has the same form as the PARAFAC1/CANDECOMP model.

To our knowledge specific investigations into the possibilities of diagonalizing the core matrix from a Tucker3 model have not been made. However, the procedures outlined for the Tucker2 model in Chapter 5 equally apply to the Tucker3 model. Most of the related work has been carried out on the IDIOSCAL model (= symmetric Tucker2 model; see section 3.3) by Cohen, (1974, 1975), MacCallum (1976b), De Leeuw & Pruzansky (1978).

When the number of components for the third mode, u, is equal to 1, the Tucker3 model becomes

$$Z_k = e_kGCH'$$

which means all Z_k are proportional, and so are the $W_k=e_kD$. In this case there always exist an A and B such that D=ACB' is diagonal, and A and B can be found from a singular value decomposition of C. If the number of components of the first two modes are equal to 2 (s=t=2), then all points $(\tilde{w}_{k1}, \tilde{w}_{k2})$ will lie on a straight line through the origin because of the proportionality of $\tilde{w}_{k1}$ and $\tilde{w}_{k2}$ as $\tilde{w}_{kp} = e_kd_p$.

Meredith (1964) and Schönemann (1972) discuss the problem of simultaneously diagonalizing two symmetric matrices by a single transformation matrix, while De Leeuw (1982) discusses the same problem in greater generality, and gives conditions under which it is possible to find such transformation matrices. In case the first and second mode are identical and u=2, the two core planes of the Tucker3 model can almost always be diagonalized. If the modes are different such diagonalization also is possible under certain conditions (De Leeuw, 1982, pers. comm.)

When u=2, $W_k = e_{k1}D_1 + e_{k2}D_2$, or $\tilde{w}_{kp} = e_{k1}d_{p1} + e_{k2}d_{p2}$, and the weights in W lie in a two-dimensional subspace of the s-dimensional space of W (i.e. the (n×s) matrix W has only rank 2). This is, of course, no restriction when s=2, but it is when s > 2, and in this sense the Tucker3 model with diagonal core matrix is a special case of the PARAFAC1/CANDECOMP model.

Summarizing, we may say that if the Tucker3 core matrix can be diagonalized (and it can be under relatively general conditions when u=1 or u=2), the Tucker3 model is equal to, or a special case of the PARAFAC1/CANDECOMP model. For most data sets we have analysed so far, one or two components for the third mode sufficed (see Table 3.2). This is partly due to the kind of centring or standardizations used, but the scalings employed are fairly typical (see Chapter 6). Thus in practice the Tucker3 model is not or hardly, more general than PARAFAC1/CANDECOMP.

Theoretically, the Tucker2 model is more general than the PARAFAC1/CANDECOMP model, especially because it contains far more parameters. Whether in practice it is also more general is not yet completely clear. In order to investigate this properly, nonsingular transformation procedures are needed to transform the extended core matrix to a diagonal form. In Chapter 5 we present such a procedure, but there are still many unsolved problems connected with this routine. Our impression so far is that in a number of applications it will be hard to find suitable transformations to transform the core matrices to diagonality. See for instance the *Attachment study* in Chapter 8 and the adjective set of the *Cola study* in Chapter 11.

Table 3.2 *Standardized weights of 'third mode' components* $

Data set	number of elements in mode	standardized component weights				total weight= SS(Fit)	type of centring£
De Gruijter	6	.92	.01			.93	ik,jk
Hohn	4	.46	.27	.13		.86	j(norm.)*
Van der Voort	5	.80	.05			.85	overall
Van de Geer	111	.80	.02	.01		.83	ik,jk
Bus	37	.70	.07			.77	jk
Osgood & Luria	6	.45	.24			.69	j
Schiffman (sim.)	10	.49	.13			.62	ik,jk
Goossens	65	.50	.09			.59	j
Lammers	11	.58	.002			.58	jk(norm.)
Lammers	11	.55	.01			.56	j(norm.)
Van der Kamp	100	.48	.04	.01		.53	ik,jk
Osgood & Luria	6	.43	.09			.52	ik,jk
Van der Kloot	65	.50	.02			.52	ik,jk
Schiffman (adj.)	10	.30	.13	.07		.50	ik,jk
Schiffman (adj.)	10	.30	.10	.07		.47	scale midpoint
Jones & Young	19	.41	.04			.45	ik,jk
De Leeuw	11	.39	.02			.41	ik,jk
Sjöberg (set 2)	100	.20	.13	.06	.01	.40	j
Miller & Nicely	17	.31	.01			.32	ik,jk

£ the predominance of ik,jk centring (see Chapter 6) is the result of the many similarity sets analysed, and an earlier tendency to ik,jk centre all data sets.

$ indicates the 'third mode', which may be considered to contain the replications or conditions.

* norm. = normalized (see section 6.2).

Although the Tucker3 model is in most cases no more general than the PARAFAC1/CANDECOMP model, it has a number of advantages over the latter. In the Tucker3 model the components in the matrices G, H, and E may be orthonormal. This means that in many cases the interpretation will be simpler for the Tucker3 model.

3.5 FACTOR ANALYSIS MODELS OR COVARIANCE STRUCTURE MODELS

Introduction. Together with his introduction of the component models for three-way data, Tucker (1966a) introduced a common factor analysis model, i.e. a model which allowed unique variances as well as common factors. His way to solve this model was analoguous to the then standard procedure for the communalities before entering into the factor analysis proper. Snyder (1968) elaborated on Tucker's idea, and included more sets of unique variances. Via an exposition of Bloxom (1968a), Bentler & Lee (1978,1979) arrived at a formulation of three-mode factor analysis models which could be handled within the framework of the general theory of the analysis of covariance structures. They also showed how the estimation procedure works, and implemented their ideas in a number of programs. Law & Snyder (1981) showed how a very general formulation of the analysis of covariance structures by McDonald (1978) includes the three-mode factor analysis model, while Bentler & Lee (1979) did the same for a similar general model of Bentler (1976). By embedding the factor analysis model in covariance structure models it becomes possible to apply the whole machinery connected with these general models. If one is prepared to make the necessary assumptions, various statistical stability statements can be made about the analysis. Moreover, it becomes easy to include restriction almost everywhere in the model, so that the testing of specific hypotheses about the structure of the solution is possible.

A general description of the models that have been proposed in this area is given below. A more detailed discussion would in fact entail an excursion into the general theory of covariance structure models, which would take us too far away from the mainstream of the present book.

It is not possible to discuss all models proposed in this area, but only some general classes. As soon as it is possible to impose restrictions on some parameters to be estimated, the concept of what constitutes a different model becomes very vague. Therefore, we will deal only with exploratory models on which no detailed constraints have been placed, for instance at the level of the individual parameter. For the random variables we will use the third mode, more in line with the individual differences approach described above than with the custom in the analysis of covariance structures.

Tucker's three-mode common factor analysis model. Tucker (1966a), p.301ff) suggested, especially when the number of individuals n is large, that it might be fruitful to look at the ij-combination mode, and analyse the correlation matrix derived from that model. Directly extending from that, he suggested treating that correlation matrix as in common factor analysis by estimating the unique variances separately, removing them from the correlation matrix, and only estimate the three-mode model from the adjusted correlation matrix. This proposal can be formulated as

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

with ε_{ijk} the *unique score* for individual k on the i-th element of the first and the j-th element of the second mode. The *correlation form* of this model becomes

$$r_{ij,i'j'} = \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} \sum_{p'=1}^{s} \sum_{q'=1}^{t} \sum_{r'=1}^{u} g_{ip} h_{jq} c_{pqr} v_{rr'} c_{p'q'r'} h_{j'q'} g_{i'p'} + u_{ij,i'j'}$$

with $v_{rr'} = \sum_{k=1}^{n} e_{kr} e_{kr'}$, the correlation between the third mode factors r and r'. Tucker took these correlations to be zero, so that $v_{rr'}$ drop out of the model. The $u_{ij,i'j'}$ are the unique variances, and they are assumed to be zero unless i=i' and j=j'. Furthermore, it has to be assumed that the third mode factors and the unique scores are uncorrelated, in order to avoid product terms in the model. Noteworthy about the model is that no direct factor loadings are available for the individuals, i.e. the e_{kr} cannot be estimated from the data. The principal reasons for this are the extra restrictions put on the e_{kr}'s and the ε_{ijk}'s. By incorporating these restrictions into the model the dimensionality of the individuals' mode is greater than it would have been without these restrictions (see Tucker, 1966a, p.310).

Snyder's unique variances model. Snyder (1968) proposed to extend Tucker's model by including unique variances for each of the first and second modes as well. The model therefore gets a rather complex form:

$$z_{ijk} = \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} g_{ip} h_{jq} e_{kr} c_{pqr} + \tilde{\varepsilon}_{ijk} + \ddot{\varepsilon}_{ijk} + \varepsilon_{ijk}$$

with a whole series of restrictions like the ones in the previous model, leading to a separable correlation form, which we will neither give nor discuss here. The usefulness of this approach is somewhat difficult to gauge as extensive studies with this model have apparently not been carried out, and the report itself never officially appeared in print. Our guess whould be that questions which necessitate the use of such a model might be better treated with some of the approaches outlined below.

Bloxom's reformulation of Tucker's model. Bloxom (1968a) seems to be the first who explicitly formulated the three-mode factor analysis model using random variables (which we will underline) instead of a finite matrix of person parameters as Tucker did. Bloxom's model takes the form

$$\underline{z}_{ij} = \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} g_{ip} h_{jq} c_{pqr} \underline{e}_{r} + \underline{\varepsilon}_{ij}$$

with $\underline{z}_{ij}$ the ij-th element of the random vector $\underline{z}$ for the ℓ×m-observations, $\underline{e}_r$ the r-th element of the u-dimensional vector $\underline{e}$ of the factor scores, and $\underline{\varepsilon}_{ij}$ the ij-th element of the vector $\underline{\varepsilon}$ of the ℓ×m-residual variates (or unique scores). In order to solve the model the (standard) assumptions are made that $\underline{e}$ and $\underline{\varepsilon}$ are statistically independent and that the residual variates $\underline{\varepsilon}$ are mutually independent. The covariance form of the model is

$$\underline{\sigma}_{ij,i'j'} = \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} \sum_{p'=1}^{s} \sum_{q'=1}^{t} \sum_{r'=1}^{u}$$

$$g_{ip} h_{jq} c_{pqr} \underline{v}_{rr'} c_{p'q'r'} h_{j'q'} g_{i'p'} + \underline{u}_{ij,i'j'}$$

with $\underline{V} = \{\underline{v}_{rr'}\} = E(\underline{e}\underline{e}')$, the covariance matrix of the factor scores, and $\underline{U} = \{\underline{u}_{ij,ij}\} = E(\underline{\varepsilon\varepsilon'})$ the matrix of unique variances of the residual variates.

A curious thing about Bloxom's presentation of his model is that although he is clearly aware of the principal component version in Tucker (1966a), he seems to neglect the factor analysis counterpart of that model in the same paper. As can be seen from the above formulas the models are effectively the same, only couched in different terms.

The Bentler & Lee models. Bentler & Lee (1979) presented some three-mode models, the most general of which is a slight extension of Bloxom's proposal

$$\underline{z}_{ij} = \underline{\mu}_{ij} + \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} g_{ip} h_{jq} c_{pqr} \underline{e}_r + \underline{\varepsilon}_{ij}$$

with $\underline{\mu}_{ij}$ the ij-th element of the random vector $\underline{\mu}$ of population means. In Bloxom's model $\underline{\mu} = \underline{0}$. The covariance form of Bentler & Lee's model is the same as that of Bloxom. It is suggested by the proposers (p.89) to impose a number of restrictions on the model. In an exploratory context these could take the form of assuming orthonormal factor scores for the individuals and requiring a special structure for G, H, and the core matrix. The purpose of all these restrictions is to make the model identifiable, and thus amenable to solving it by generalized least squares or maximum likelihood methods. Implicit in this procedure is, that the structure imposed by these restrictions is one that make sense within the context of the data. If a model can be found that fits the data well, then one possible model has been found, but, only within the limits set by the earlier restrictions. With the large number of parameters under consideration other structures induced by other sets of restrictions could have given entirely different sets of models.

A special model, also treated in an earlier paper (Bentler & Lee, 1978), has the same form as the model above but has the additional restrictions that S = $\{s_{pq,p'q'}\}$ is an identity matrix, thus

$$\underline{s}_{pq,p'q'} = \sum_{r=1}^{u} \sum_{r'=1}^{u} c_{pqr} \underline{v}_{rr'} c_{p'q'r'} \begin{array}{ll} = 1 & \text{if } p=p' \text{ and } q=q' \\ = 0 & \text{otherwise} \end{array}$$

These restrictions are introduced without much explanation or justification and without an indication of the consequences for the interpretation of the model.The same model was proposed by Lohmöller (1978, p.4), who called $S=\{s_{pq,p'q'}\}$, the 'core covariance matrix'. He indicated what the implications are when this matrix is equated to an identity matrix. In section 13.3 we give an extensive discussion of this *latent covariation matrix*, as we prefer to call it.

The covariance form of the model, including the restrictions

$$\underline{\sigma}_{ij,i'j'}=\sum_{p=1}^{s}\sum_{q=1}^{t} g_{ip}h_{jq}g_{i'p}h_{j'q}+\underline{u}_{ij,i'j'}$$

looks like a factor analysis analogue of the PARAFAC1 model with a random third mode (or E orthonormal), and an unequal number of components in the first two modes:

$$z_{ijk}=\sum_{p=1}^{s}\sum_{q=1}^{t}\sum_{r=1}^{u} g_{ip}h_{jq}e_{kr}$$

which has as its covariance form under the assumption of orthonormal components e_r

$$\sigma_{ij,i'j'}=\sum_{p=1}^{s}\sum_{q=1}^{t} g_{ip}h_{jq}g_{i'p}h_{j'q}$$

In this way the specialized Bentler & Lee model has an interpretation analoguous to PARAFAC1 in the sense of proportional profiles (see also Harshman, 1972, p.32-34; Harshman & Berenbaum, 1981).

Three-mode factor analysis as a covariance structure model. Both Bentler (1976) and McDonald (1978) have proposed very general models which probably subsume almost all conceivable models to analyse covariance matrices. Bentler & Lee (1979) show how the three-mode factor analysis model can be placed in Bentler's general model, while Law & Snyder (1981) show how the same model fits into McDonald's model. The attractive aspect of this embedding in general models is that programs to solve the general model can be used to attack three-mode factor analysis as well. However, as Bentler & Lee point out, more specialized programs can handle the specific models more efficiently, and produce output which is more geared towards the special application. Bentler & Lee (1979) also show that their (1978) model can be seen as a special case of Jöreskog's (1973a) second-order factor model (see also Bloxom, forthcoming), so that a program like LISREL (Jöreskog & Sörbom, 1978) can solve this particular problem as well (cf. section 2.2).

3.6 INDIVIDUAL DIFFERENCES MODELS OR COVARIANCE STRUCTURE MODELS?

Bentler & Lee (1979, p.79) contrast the two types of models by saying: "Tucker's development of his model is primarily consistent with a traditional approach towards data analysis rather than a statistical approach towards parameter estimation and model evaluation". They continue to say that the standard machinery of statistical evaluation of a model (standard errors for the estimators, goodness-of-fit tests, etc.) is not available for Tucker's approach, and, in addition, the properties of the estimators are not known.

Within the framework of the analysis of covariance structures, using maximum likelihood or generalized least squares estimation, evaluation is possible, and the properties of the estimators are known, at least asymptotically. All these advantages are bought by introducing (asymptotic) assumptions about the data and/or error structure, and by imposing a number of largely arbitrary restrictions on the model, especially in the exploratory context. Such restrictions are necessary to identify the model and to solve the estimation problem. The restrictions entail an a priori choice in favour of certain classes or kinds of models. With respect to the restrictions, Bentler & Lee (1979, p.99) admit that in an exploratory context the identification constraints can be viewed as temporary and arbitrary, and should be removed by appropriate transformations to meaningful solutions in the dimensional space. This suggestion has the flavour of a proposal to use INDSCAL because of its uniqueness property, and rotating afterwards because the stimulus space is not really what was desired. The strength of the covariance structure approach lies in the confirmatory context using (very?) large samples. In such situations it is possible to impose restrictions, which have substantive implications on the parameters that matter. In addition, multivariable-multicondition matrices can be analysed with covariance structure models, and not efficiently by individual differences models, as the latter can only be used for within-condition covariances matrices and not for between-conditions covariances.

The Alternating Least Squares (ALS) approach towards solving the three-mode principal component model which forms the foundation of this book has some of the defects, but also some of the advantages of both approaches. We will treat most of the details in the chapters to come, and only point out some of the properties here.

One of the problems Bentler & Lee quote for Tucker's approach is the unknown nature of the estimators of the parameters. However, although the estimators are not least squares ones and their properties not known, the Tucker methods are sufficiently simple so that, in principle, it should be possible to work out some statistical stability statements, especially for the component loadings. In the alternating least squares approach, in which sets of parameters are estimated iteratively by conditional least squares procedures, the statistical stability of the estimators can be determined in principle, but the method is much more complex than Tucker's methods, and the task to determine the statistical stability will be much more difficult. The properties of the estimators, of course, also depend very much on the distributional assumptions one is prepared to make, and specifying the appropriate assumptions for three-mode data is not an easy task. Assessing the fit of a model is possible within the covariance structure framework, but not for Tucker's approach.

In section 4.3 it is shown that for the alternating least squares procedures the total sum of squares of the data can be partitioned up to the level of the individual data points making assessment of the fit of the model possible, be it still in an exploratory way.

Finally, using a theorem of Kruskal (Carroll, Pruzansky & Kruskal, 1980), it is possible to show that the ALS framework can be used to incorporate restrictions on the configurations or components (see section 3.7). The details of incorporating such restrictions for three-mode principal component analysis still have to be worked out, but Lohmöller (1981b), for instance, has developed alternating least squares (or partial least squares, as he calls it) procedures which can handle succesfully all or most of the models the standard programs for the analysis of covariance structures, like LISREL (Jöreskog & Sörbom, 1978) can handle. (For a comprehensive comparison, see Dijkstra, 1981). In other words the ALS approach serves most needs of exploratory approach, and has some confirmatory capabilities as well.

3.7 EXTENSIONS OF THREE-MODE PRINCIPAL COMPONENT MODELS

Missing data. In the present formulation of the Tucker3 and Tucker2 models no option has been provided for the inclusion of missing data. In all examples in this book the starting point is a completely crossed design with complete information. The alternating least squares approach to solving the estimation problem of the model makes it relatively simple to include facilities for handling missing data.

All that is necessary is to include in the algorithm an extra phase in which the missing data points are estimated after each step of the main algorithm. The missing data points are initialized by an arbitrary or informed estimate, and the algorithm then estimates the parameters of the model on the basis of these augmented data. Using these parameter estimates the missing data points are re-estimated in a regression-like fashion; then a new cycle of the iteration is started. In fact, such procedures are standard within the ALS-approach to most metric models (see, for instance, Young, De Leeuw, & Takane, 1980; Young, 1981).

Extensions to other measurement characteristics. A further application of the same principle is to include a so-called optimal scaling phase in the ALS algorithm to accommodate data with various measurement characteristics. Young and De Leeuw, and their co-work-ers (e.g. Young, De Leeuw, & Takane, 1980; Young, 1981) distinguish three measurement characteristics to classify the types of data commonly found in social science research, viz. measurement process, measurement level, and measurement conditionality. The first of the three refers to the discreteness or continuity of the observed values, in other words, whether the observed numbers represent intervals or intrinsically discrete values. As for measurement levels one generally distinguishes between the nominal, ordinal, interval, and ratio levels. Finally, the measurement conditionality of data indicates which data points may be compared. For instance, in some applications, only comparison within elements of the third mode (i.e. within frontal planes) may be made but no comparisons can be meaningfully made across frontal planes without further scaling. A detailed discussion of these measurement characteristics can be found in the above references.

The first inclusion of an optimal scaling phase in three-mode data was given by Takane, Young, & De Leeuw (1977) in their paper on scaling individual differences using ALSCAL, which treats a large number of the intricacies connected with this inclusion. Sands & Young (1980) discuss in a similar vein the combination of an optimal scaling phase with both the CANDECOMP-PARAFAC1 model (called "weighted model") and the "replicated-model" in which subjects are treated as true replications. From these two papers it is evident that a similar inclusion of optimal scaling into the TUCKALS algorithms is feasible. The inclusion or imputation of missing data can be incorporated in a natural way in the optimal scaling phase.

External analysis. In the TUCKALS algorithms all component matrices are estimated, i.e. G, H, and E for the Tucker3 model, and G and H for the Tucker2 model. It is, however, also feasible to require that one or two of G, H, and E in the Tucker3 model, or G or H in the Tucker2 model remain fixed, while the other component matrix (matrices) are estimated via the ALS-procedure. Such an 'external' analysis was proposed by Carroll (1972) in connection with unfolding. It is included in ALSCAL-4 (Young & Lewyckyj, 1979), and in the ALSCOMP3 program devised by Sands & Young (1980).

In section 11.4 the usefulness of such an external analysis using the Tucker2 model is illustrated in principle, but not in practice, as the procedure is not yet included in the TUCKALS programs.

Another possible application comes from an extension of the research reported in part in Chapter 10, and in full in Van der Kloot & Kroonenberg (1982). In the new experiment eleven hypothetical stimulus persons characterized each by one personality trait descriptor were rated by 95 subjects on all of the same eleven descriptors, or scales. The derived scale configuration (from an 11×11×95 TUCKALS2 analysis) thus indicates how the scales are related. In addition to the eleven single-descriptor stimulus persons, each subject also rated a number of two-, three-, four-, and/or five-descriptor stimulus persons, but not all of these stimulus persons were rated by all the subjects.

If we assume that the scale space derived from the single-description stimulus persons reflects the way the scales are organized, we can use this scale space to determine the way the subjects arrange the more-descriptor stimulus persons in a coherent fashion. By performing external analysis for the subjects who rated the same more-descriptor persons, keeping the scale space fixed, it is possible to compare the groups which rated different combinations of more-descriptor persons. The full analysis and more detailed considerations of external analysis with three-mode principal component analysis can be found in Kroonenberg & Van der Kloot (Note 3), while a somewhat different approach to external analysis of the same data can be found in Van der Kloot, Bakker, & Kroonenberg (Note 4).

Restrictions on configurations. In Chapter 5 we discuss transformations to obtain a simpler structure of the Tucker2 core matrix. Following the practice in standard principal component and common factor analysis, it is also possible to transform the component matrices by some orthogonal transformation procedure such as Kaiser's (1958) varimax procedure, or by a non-singular transformation. Both these procedures aim at obtaining some kind of 'simple structure' in an exploratory fashion. In situations where there is considerable knowledge about the underlying theory, however, it seems more appropriate to use some kind of confirmatory approach by specifying restrictions on the parameters of the model, although even in an exploratory context, methods to impose restrictions can be useful in finding simple structures (see Jöreskog, 1978; Kroonenberg & Lewis, 1982).

The various models for treating three-mode data (as outlined in sections 3.3 to 3.6) were initially viewed as a series of progressively more specialized models. Possibly inspired by the development of covariance structure models, the three-mode models came to be seen as models which could be sequentially tested by increasing or decreasing the number of constraints on the parameters. In fact, Bentler & Lee (1978, 1979), and Lohmöller & Wold (1980) indicate ways to do this for their restricted models. Especially in multidimensional scaling, minimization of loss functions under constraints using some type of alternating least squares procedure has been eminently succesful (Bentler & Weeks, 1978; Bloxom, 1978; Borg & Lingoes, 1980; De Leeuw & Heiser, 1980; Carroll, Pruzansky, & Kruskal, 1980). The latter prove a theorem which shows that solving the three-mode principal component model under constraints is possible (p. 10). The approach using constraints seems a more appropriate way to assess the adequacy of a model compared to a more general or a more restricted model than the approaches outlined in Chapter 5, and those given by MacCallum (1976a,b), Cohen (1974,1975), and De Leeuw & Pruzansky (1978), which attempt to transform the core matrix to a specified target.

3.8 THREE-MODE CAUSAL MODELLING

Recent years have seen an upsurge of the use of so-called 'causal modelling'. Elaborate stochastic models are defined which describe causal networks in terms of latent variables *(structural model)* coupled with an explicit formulation of the relations between between the observed and latent variables *(measurement model)*. Especially Jöreskog (e.g. 1973), Bentler (e.g. 1976), Mc Donald (e.g. 1978), and Wold (e.g. 1975) have proposed both general models as well as developed algorithms and programs to solve the parameter estimation of these models. The main data to analyse are the variances and covariances of the observed variables, sometimes in conjunction with the means.

The three-mode models specified by Bloxom (1968a), Bentler & Lee (1978,1979), and Lee & Fong (1982) belong to this class (as outlined in section 3.6), although they are only measurement models. Lohmöller & Wold (1980) propose, and use a somewhat less elaborate measurement model in conjunction with a structural model, calling their proposal a *three-mode path model:*

$$\underline{z}_{ij} = \sum_{p=1}^{s} \sum_{q=1}^{t} g_{ip} h_{jq} \underline{\eta}_{pq} + \underline{\varepsilon}_{ij} \qquad \text{measurement model (=Bentler \& Lee, 1978)}$$

$$\underline{\eta}_{pq} = \sum_{p'=1}^{s} \sum_{q'=1}^{t} a_{pp'} b_{qq'} \underline{\eta}_{p'q'} + \underline{\delta}_{pq} \qquad \text{structural model}$$

They augment their model by including for each parameter matrix a 'pattern matrix' (design matrix is probably a better word) consisting of zeroes and ones which specifies the parameters to be estimated:

$$\underline{z}_{ij} = \sum_{p=1}^{s} \sum_{q=1}^{t} (\mu_{ip} g_{ip}) (\nu_{jq} h_{jq}) \underline{\eta}_{pq} + \underline{\varepsilon}_{ij}$$

$$\underline{\eta}_{pq} = \sum_{p'=1}^{s} \sum_{q'=1}^{t} (\alpha_{pp'} a_{pp'}) (\beta_{qq'} b_{qq'}) \underline{\eta}_{p'q'} + \underline{\delta}_{pq}$$

with μ_{ip}, ν_{jq}, $\alpha_{pp'}$, and $\beta_{qq'}$ zero or one, according to the design specifications. In addition, restrictions may be imposed on the covariation (covariance or correlation) matrix of the $\underline{\eta}$.

In their report, Lohmöller & Wold outline an ALS (or PLS) algorithm to solve the three-mode path model, but few details are given with respect to the performance of the algorithm. However, a program description is announced.

As mentioned before, the measurement model is somewhat less elaborate in the sense that no core matrix is specified explicitly. However, the covariation (covariance or correlation matrix) $\underline{S}$ = $E(\underline{\eta}\underline{\eta}')$ is the component analogue of the variance-covariance matrix of the random vector of observations $E(\underline{z}\underline{z}')$. If we conceive of the entries in the Tucker3 core matrix as the scores of idealized subjects on latent variables under prototype conditions (see section 2.2 and section 6.9), then $\underline{S}$ may be viewed as the covariation matrix of those scores, and will be referred to as the *latent covariation matrix* (see section 13.3 for a further discussion of this matrix).

A final point, taken up again in Chapter 13, is that Lohmöller & Wold (1980) use a multivariate autoregressive model as their basic structural model, with which they analyse longitudinal data.

3.9 *n*-MODE EXTENSIONS

A number of extensions to *n*-mode individual differences scaling and principal component analysis have been proposed. Carroll & Chang (1970), Harshman (1970), and Carroll & Wish (1974) mention extensions of the INDSCAL and PARAFAC1 models (see sections 3.2, and 3.3), Carroll & Chang (1970) have implemented their proposal in their CANDECOMP program for *n* up to 7 (p.313). The only applications known to us of this model are Green, Carmone, & Wachspress (1976), and Carroll, Pruzansky, & Green (1977).

Lastovicka (1981) discusses a direct four-mode extension of Tucker's (1966a) Method I by defining a standard principal component analysis for the fourth mode, analoguous to the already available ones for the other three modes. Similarly the TUCKALS algorithms can be extended by adding more substeps (see section 4.4). In principle, the extension to *n*-mode data is straightforward, but it becomes increasingly more complex to keep track of the summations over the proper indices. Moreover, the description of such an *n*-mode procedure becomes exceedingly cumbersome without new notation.

Such a new notation based on permutation matrices was devised by Kapteijn, Neudecker, & Wansbeek (1982) who describe what amounts to a Tucker*n* model, and they describe a TUCKALS*n* algorithm. They also propose to use Tucker's Method I to initialize the algorithm just as we have done for TUCKALS algorithms.

The number of published applications with n greater than three is very small. Primarily, because of the lack of programs to perform the analyses, but also because of the complexity of the interpretation. Furthermore, it is quite a task to generate the appropriate data in sufficient quantities to make such an analysis meaningful. It is no problem to conceptualize adequate data, after all Cattell's (1966a) data box has ten ways, but few investigators seem to have taken the trouble to collect such multi-mode data, and try to look at what are essentially four-mode and higher interactions.

METHODS AND ALGORITHMS

4

G-substep

$$p^{a}_{ii'} = \sum_{j=1}^{m} \sum_{j'=1}^{m} \sum_{k=1}^{n} \sum_{k'=1}^{n} \sum_{q=1}^{t} \sum_{r=1}^{u} h^{a}_{jq} h^{a}_{j'q} e^{a}_{kr} e^{a}_{k'r} z_{ijk} z_{i'j'k'}$$

$$G_{a+1} = P_a G_a (G'_a P^2_a G_a)^{-\frac{1}{2}}$$

H-substep

$$q^{a}_{jj'} = \sum_{k=1}^{n} \sum_{k'=1}^{n} \sum_{i=1}^{l} \sum_{i'=1}^{l} \sum_{r=1}^{u} \sum_{p=1}^{s} e^{a}_{kr} e^{a}_{k'r} g^{a+1}_{ip} g^{a+1}_{i'p} z_{ijk} z_{i'j'k'}$$

$$H_{a+1} = Q_a H_a (H'_a Q^2_a H_a)^{-\frac{1}{2}}$$

E-substep

$$r^{a}_{kk'} = \sum_{i=1}^{l} \sum_{i'=1}^{l} \sum_{j=1}^{m} \sum_{j'=1}^{m} \sum_{p=1}^{s} \sum_{q=1}^{t} g^{a+1}_{ip} g^{a+1}_{i'p} h^{a+1}_{jq} h^{a+1}_{j'q} z_{ijk} z_{i'j'k'}$$

$$E_{a+1} = R_a E_a (E'_a R^2_a E_a)^{-\frac{1}{2}}$$

4.1 INTRODUCTION

In the previous chapter we have reviewed a large number of scalar-product models for three-mode data. In this chapter we will concentrate on two of them, viz. the Tucker3 and Tucker2 models. Both are the most general in the class of component models with three and two reduced modes, respectively. As much of the development of the solutions for the two Tucker models runs almost parallel, we will treat only the Tucker3 model in detail, and the Tucker2 model very briefly. After a discussion of the widely used methods developed by Tucker (1966a) to solve the three-mode principal component model and of their advantages and disadvantages, we will present in some (theoretical) detail an alternating least squares procedure for the Tucker3 model. This presentation will include theorems about the conditions for a unique solution, the development of an algorithm, and convergence properties of the algorithm. Furthermore two small examples using three-mode Hilbert matrices will be used to assess the accuracy of the algorithm, and a small Monte Carlo study will be reported which was designed to assess how increasing errors influence the recovery of a particular component space.

4.2 TUCKER'S METHODS FOR THREE-MODE PRINCIPAL COMPONENT ANALYSIS

Tucker (1966a, p.294-301) describes three methods to deal with the estimation of the parameters in the three-mode principal component model. The last of these models belongs more to the covariance structure approach, and will not be discussed here.

The essence of *Method I* is that cross-product matrices P, Q, R are formed for the three modes:

$$p_{ii'} = \sum_{j=1}^{m} \sum_{k=1}^{n} z_{ijk} z_{i'jk} \; ; q_{jj'} = \sum_{k=1}^{n} \sum_{i=1}^{\ell} z_{ijk} z_{ij'k}$$

$$r_{kk'} = \sum_{i=1}^{\ell} \sum_{j=1}^{m} z_{ijk} z_{ijk'}$$

and that standard principal component matrices, say G, H, and E, are computed via an eigenvalue-eigenvector decomposition for each of them. In general one only wants to retain the first few eigenvectors, assuming that the other vectors correspond to random variation in the data rather than systematic variation. G, H, and E are thus columnwise orthonormal, rather than orthonormal. From the formulation of the model it can be derived, as Tucker does for his method (p.292) and we will do for ours below in section 4.3, that the matrices G, H, and E are all that is necessary to arrive at an estimate for the parameters of the core matrix:

$$c_{pqr} = \sum_{i=1}^{\ell} \sum_{j=1}^{m} \sum_{k=1}^{n} g_{ip} h_{jq} e_{kr} z_{ijk}$$

As long as all eigenvectors corresponding to non-zero roots are retained the estimators for c_{pqr} turn out to be least squares ones. However, when G, H, and E are truncated to the eigenvectors corresponding to the s, t, and u largest eigenvalues, then the estimators of the core matrix are no longer least squares ones, as the deletions of the later eigenvectors in the component matrices affect the estimators of the core matrix in a complicated way. This implies, for instance, that the relation

$$\sum_{i=1}^{\ell} \sum_{j=1}^{m} \sum_{k=1}^{n} z^2_{ijk} = \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} c^2_{pqr}$$

which is true for the complete decomposition with $s=\ell$, $t=m$, and $r=n$, no longer has a simple equivalence for the reduced decomposition, e.g. in the form of a separate sum of squares for the core matrix and one for the error. In section 4.3 and Chapter 7 it will be shown, that the partioning of the total sum of squares is a very powerful tool for interpretation. With Tucker's Method I it is not clear how much of the total sum of squares of the data is accounted for by the reduced decomposition. Each of the modes will nearly always have different amounts of variation accounted for. It should, however, be noted that the better a model fits, the smaller the differences in amount of variation accounted for between the modes will be.

In the complete decomposition each squared element, c^2_{pqr}, of the core matrix indicates the amount of variation accounted for by that element, or the combination of components that this element stands for - here the p-th component of the first mode, the q-th component of the second mode, and the r-th component of the third mode. This interpretation cannot be maintained when the amount of variation accounted for is different for each mode.

Tucker's *Method II* differs from Method I in the sequence of calculations. The components of the first mode are computed first, and they are immediately used to reduce the total data matrix to dimensions s×m×n. This reduced matrix is then employed for the computation of the component matrix for the second mode, and the resulting components are used to reduce the data matrix once again now to dimension s×t×n. A final singular value decomposition (see section 2.2) is used to find the core matrix, and the remaining components for the third mode. The purpose of this procedure is to circumvent the solving of the possibly large eigenvalue-eigenvector problem for one of the modes (usually individuals). The problems with discarding small roots are rather serious in this procedure, as the errors of approximation at one stage are passed on to the next stage. Incidentally, this way of solving the Tucker3 model demonstrates Bloxom's (forthcoming) point that the model is a third-order factor component model, and it fits nicely in the presentation of the model as three nested sets of linear combinations (see section 2.2).

Summing up it seems fair to say that, although Tucker's methods are acceptable in case of a complete decomposition, and in case of a good fit, they introduce problems in the reduced case with respect to the properties of the estimators, the interpretation of the core matrix, and the amount of variation accounted for by the components.

In the following sections we develop procedures which do not suffer from these defects, but which are conceptually and computationally more complex. It will be possible with the results of these procedures to assess the quality of the fit of the model and to partition the total sum of squares (see Chapter 7).

4.3 LEAST SQUARES SOLUTIONS FOR THE TUCKER3 MODEL

Loss functions. In section 2.3 and 3.2 we introduced the Tucker3 model

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

with the restrictions that the columns of $G = \{g_{ip}\}$, $H = \{h_{jq}\}$, and $E = \{e_{kr}\}$ are orthonormal. A matrix formulation of the model can be given if we use Kronecker products, $\otimes$, of the components (Bellman, 1960), and 'combination modes' (Tucker, 1966a, p.289) for the data matrix and the core matrix:

$$Z = GC(H' \otimes E')$$

with the $(\ell \times mn)$ matrix Z, and the $(s \times tu)$ matrix C written with combination modes. By symmetry other matrix formulations are possible by using different combination modes and other component matrices in the Kronecker product. By using summation notation we can avoid the use of Kronecker products. In doing so, one formulation will suffice, and the symmetry of the model in the three modes is reflected by the formulation.

If we were interested in exactly decomposing Z into all its components, Tucker's methods would suffice to provide a solution for the decomposition, as we remarked above. However, in practical applications one is interested only in the first few principal components for each of the modes. In general this precludes finding an exact decomposition of Z into G, H, E, and C. One therefore has to be satisfied with an approximation $\tilde{Z} = GC(H' \otimes E')$, i.e. finding G, H, E, and C such that the difference between the model and the data is minimal according to some loss function. In slightly different terms, we have to look for the best approximate decomposition $\tilde{Z}$ of the three-mode matrix into G, H, E, and C according to the Tucker3 model.

In our case, as in many similar situations, we define a mean-squared loss function. Thus we search for an approximate solution $\tilde{Z}$ such that

$$f(G,H,E,C) = [[Z-\tilde{Z}]]^2 = \sum_{i=1}^{\ell} \sum_{j=1}^{m} \sum_{k=1}^{n} (z_{ijk}-\tilde{z}_{ijk})^2 =$$

$$= \sum_{i=1}^{\ell} \sum_{j=1}^{m} \sum_{k=1}^{n} (z_{ijk} - \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} g_{ip}h_{jq}e_{kr}c_{pqr})^2$$

is minimal. The minimization has to be carried out under the restrictions of the model, i.e. G, H, and E have to be columnwise orthonormal. The Z for which f attains its minimum will be designated as $\hat{Z} = \{\hat{z}_{ijk}\}$, with

$$\hat{z}_{ijk} = \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} \hat{g}_{ip}\hat{h}_{jq}\hat{e}_{kr}\hat{c}_{pqr}$$

Existence of a best approximate solution. Our first task is to show that there exist indeed such G, H, E, and C that the loss function attains a (global) minimum. This can be done by showing first that for given G, H, and E a unique C can be found which minimizes the loss function, and which can be expressed in terms of G, H, and E. In this way we are left with a minimization over G, H, and E.

A lemma due to Penrose (1955 - see also Kroonenberg & De Leeuw, 1977, p.3-3, 3-4) states that there exists a unique $\hat{C}$, such that the function h,

$$h(C) = [[Z-\tilde{Z}]]^2 = [[Z-GCF']]^2$$

is as small as possible. This $\hat{C}$ is equal to G'ZF, and the absolute minimum zero is reached only if Z = GG'ZFF'. By defining the (mn×qr) matrix $F = \{f_{jk,qr}\}$ with $f_{jk,qr} = h_{jq}e_{kr}$, we may conclude that $\hat{C} = \{\hat{c}_{pqr}\}$

$$c_{pqr} = \sum_{i'=1}^{\ell} \sum_{j'=1}^{m} \sum_{k'=1}^{n} g_{i'p}h_{j'q}e_{k'r}z_{i'j'k'}$$

minimizes f for fixed G, H, and E, and that

$f(Z) = 0$ iff $z_{ijk} =$

$$= \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} \sum_{i'=1}^{\ell} \sum_{j'=1}^{m} \sum_{k'=1}^{n} g_{ip} g_{i'p} h_{jq} h_{j'q} e_{kr} e_{k'r} z_{i'j'k'} .$$

The implication of this result is that we may minimize f over G, H, and E, after substituting the expression for $\hat{C}$ into the loss function. The optimal core matrix $\hat{C}$ can then be computed afterwards from the optimal $\hat{G}$, $\hat{H}$, and $\hat{E}$ by substituting their values in the above equation for $\hat{C}$.

The function f thus has to be optimized over the domain S

$S = \{s \mid s = (G,H,E),\ G,H, \text{ and } E \text{ columnwise orthonormal}\}$

which is a compact subset in a finite dimensional real space. Furthermore, f is a bounded continuous function on S, therefore there exists a point s = (G,H,E) in S, such that f attains its minimum. In other words the minimization problem always has a solution.

Partitioning of total sum of squares. Using the optimal $\hat{C}$ from the previous section we can thus write $\tilde{Z}$ as

$$\tilde{z}_{ijk} = \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} \sum_{i'=1}^{\ell} \sum_{j'=1}^{m} \sum_{k'=1}^{n} g_{ip} g_{i'p} h_{jq} h_{j'q} e_{kr} e_{k'r} z_{i'j'k'} .$$

Using this formulation we can rewrite the loss function as

$$f'(G,H,E) = \sum_{i=1}^{\ell} \sum_{j=1}^{m} \sum_{k=1}^{n} (z_{ijk} - \tilde{z}_{ijk})^2 =$$

$$= \sum_{i,j,k} z^2_{ijk} + \sum_{i,j,k} \tilde{z}^2_{ijk} - 2 \sum_{i,j,k} z_{ijk} \tilde{z}_{ijk}$$

The last term can be shown to be equal after algebraic manipulation, using the orthonormality of G, H, and E, - to

$$2 \sum_{i,j,k} \sum_{p,q,r} \sum_{i'j'k'} g_{ip} g_{i'p} h_{jq} h_{j'q} e_{kr} e_{k'r} z_{ijk} z_{i'j'k'} = 2 \sum_{i,j,k} \tilde{z}_{ijk}^{2}$$

In Appendix 4.1 we show that the orthonormality is not necessary for the above equation to hold but as we already assumed that G, H, and E are orthonormal, it is convenient to use this property. In a sense demanding orthonormality in a fixed model is thus analoguous to assuming statistical independence between the model and the error vectors in models with one random vector (see the discussion in section 3.5)

Thus

$$f'(G,H,E)) = \sum_{i=1}^{\ell} \sum_{j=1}^{m} \sum_{k=1}^{n} (z_{ijk} - \tilde{z}_{ijk})^2 = \sum_{i,j,k} z^2_{ijk} - \sum_{i,j,k} \tilde{z}^2_{ijk}$$

A convenient way of expressing this is as follows

SS(Residuals) = SS(Data) - SS(Model), or
SS(Res) = SS(Total) - SS(Fit).

Thus the total sum of squares may be partitioned into a residual sum of squares and a fitted sum of squares, and the above formulation shows that the minimization of the SS(Res) is equal to the maximization of the SS(Fit).

It is furthermore worthwhile to note that both the SS(Fit) and the SS(Res) can be further partitioned in many different ways, none of which involve cross-product terms. The sum of squares are after all built up from the squares of the contributions of each point (i,j,k) to the SS(Fit) and the SS(Res). We may therefore partition them in an analysis-of-variance way in order to find influential or ill-fitting elements in each of the modes, or even three-mode combinations which contribute too much or too little to the fit or residual. These matters will be taken up in Chapter 7.

Nature of the approximate solution. As mentioned in the previous section we can either minimize the residual sum of squares, or maximize the fitted sum of squares to find Ĝ, Ĥ, and Ê. It turns out to be more efficient to maximize the SS(Fit):

$$p(G,H,E) = SS(Fit) = \sum_{i=1}^{\ell} \sum_{j=1}^{m} \sum_{k=1}^{n} \tilde{z}^2_{ijk}$$

We have already shown that

$$\sum_{i,j,k} \tilde{z}^2_{ijk} = \sum_{i,j,k} z_{ijk} \tilde{z}_{ijk} \text{ , and}$$

thus

$$p(G,H,E) = \sum_{i,j,k} \sum_{p,q,r} \sum_{i',j',k'} g_{ip} g_{i'p} h_{jq} h_{j'q} e_{kr} e_{k'r} z_{ijk} z_{i'j'k} \text{ ,}$$

with the constraints that G, H, and E are columnwise orthonormal.

Incorporating these constraints into the maximization equation we get

$$p'(G,H,E,\Lambda,M,N)= \sum_{i,j,k} \sum_{p,q,r} \sum_{i',j',k'} g_{ip}g_{i'p}h_{jq}h_{j'q}\, e_{kr}e_{k'r}z_{ijk}z_{i'j'k'}$$

$$- \sum_{p} \sum_{p'} \sum_{i} \lambda_{pp'}(g_{ip}g_{ip'} - \delta^{pp'}) - \sum_{q} \sum_{q'} \sum_{j} \mu_{qq'}(h_{jq}h_{jq'} - \delta^{qq'})$$

$$\sum_{r} \sum_{r'} \sum_{k} \nu_{rr'}(e_{kr}e_{kr'} - \delta^{rr'})$$

which reads in matrix notation

$$p'(G,H,E,\Lambda,M,N)=\mathrm{tr}G' \{Z(H\otimes E)(H'\otimes E')Z'\} G - \mathrm{tr}\ \Lambda(G'G-I_s)$$

$$- \mathrm{tr}\ M(H'H-I_t) - \mathrm{tr}\ N(E'E-I_u)$$

The maximum of p follows from the requirement that the first order partial derivatives of p' are simultaneously zero at the maximum of p, and the Hessian is negative. In *Theorem 4.1* the nature of the solution of the maximization is given, while the proof can be found in the Appendix to Kroonenberg & De Leeuw (1980).

Theorem 4.1 (Approximate solution)

Let the following quantities be defined:

- Z is a three-mode data matrix
- $p = \mathrm{tr}\ G'\{Z(HH'\otimes EE')Z'\}\ G$
- $S = \{s \mid s = (G,H,E)\ G,H,E \text{ columnwise orthonormal}\}$
- U is eigenvector matrix of $P = \{p_{ii'}\}$,

$$p_{ii'} = \sum_{j=1}^{m} \sum_{j'=1}^{m} \sum_{k=1}^{n} \sum_{k'=1}^{n} \sum_{q=1}^{t} \sum_{q'=1}^{u} h_{jq}h_{j'q}e_{kr}e_{k'r}z_{ijk}z_{i'j'k'}$$

- V is eigenvector matrix of $Q = \{q_{jj'}\}$,

$$q_{jj'} = \sum_{k=1}^{n} \sum_{k'=1}^{n} \sum_{i=1}^{\ell} \sum_{i'=1}^{\ell} \sum_{r=1}^{u} \sum_{p=1}^{s} e_{kr}e_{k'r}g_{ip}g_{i'p}z_{ijk}z_{i'j'k'} ,$$

- W is eigenvector matrix of $R = \{r_{kk'}\}$,

$$r_{kk'} = \sum_{i=1}^{\ell} \sum_{i'=1}^{\ell} \sum_{j=1}^{m} \sum_{j'=1}^{m} \sum_{p=1}^{s} \sum_{q=1}^{t} g_{ip}g_{i'p}h_{jq}h_{j'q}z_{ijk}z_{i'j'k'} ,$$

- $(U,V,W)\varepsilon S$.

Then

a. $(\hat{G},\hat{H},\hat{E})\varepsilon S$ is a stationary point of p if and only if $\hat{G} = U$, $\hat{H} = V$, and $\hat{E} = W$, or orthonormal rotations thereof;

b. $(\hat{G},\hat{H},\hat{E})\varepsilon S$ maximizes p if and only if their columns are eigenvectors corresponding to the largest s,t, and u eigenvalues of $P(\hat{H},\hat{E})$, $Q(\hat{E},\hat{G})$, and $R(\hat{G},\hat{H})$ respectively, or orthonormal rotations thereof.

▼▼

In essence this theorem tells us that p can be maximized by simultaneously solving the eigenvalue-eigenvector problems of P, Q, and R. Not surprisingly this cannot be done analytically, but only iteratively.

Nature of exact solution. Apart from knowing what the approximate solution looks like, it is also of theoretical and practical importance to know what the exact solution of the maximization problem looks like, if only the assess how well we have succeeded with our approximation. We will see that for the initialization of the algorithm for the approximate solution, the outcome of the theorem is relevant as well (section 4.5). As before we will only state the result here, the proof can be found in the Appendix to Kroonenberg & De Leeuw (1980).

Theorem 4.2 (Exact solution)

A:

Let the following quantities be defined:

- Z is a three-mode matrix

- $f(G,H,E,C) = \sum_{i=1}^{\ell} \sum_{j=1}^{m} \sum_{k=1}^{n} (z_{ijk} - \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} g_{ip}h_{jq}e_{kr}c_{pqr})^2$

- $p(G,H,E) = \sum_{i} \sum_{j} \sum_{k} \sum_{p} \sum_{q} \sum_{r} \sum_{i'} \sum_{j'} \sum_{k'} g_{ip}g_{i'p}h_{jq}h_{j'q}e_{kr}e_{k'r}z_{ijk}z_{i'j'k'}$

- $(G,H,E)\varepsilon S$

- $C = \{c_{pqr}\}$, with $c_{pqr} = \sum_{i'=1}^{\ell} \sum_{j'=1}^{m} \sum_{k'=1}^{n} g_{i'p}h_{j'q}e_{k'r}$

Then the following statements are equivalent:

1. $f(\hat{G},\hat{H},\hat{E},\hat{C}) = 0$,

2. $p(\hat{G},\hat{H},\hat{E}) = \sum_{i=1}^{\ell} \sum_{j=1}^{m} \sum_{k=1}^{n} z_{ijk}^2$

3. $z_{ijk} = \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} \sum_{i'=1}^{\ell} \sum_{j'=1}^{m} \sum_{k'=1}^{n} \hat{g}_{ip}\hat{g}_{i'p}\hat{h}_{jq}\hat{h}_{j'q}\hat{e}_{kr}\hat{e}_{k'r}z_{i'j'k'}$

4. $(\hat{G},\hat{H},\hat{E},\hat{C})$ is an exact solution of the minimization problem.

B.1:

Let $(\hat{G},\hat{H},\hat{E},\hat{C})$ be an exact solution of the minimization problem, Then

$\hat{G}$ is the eigenvector matrix (or an orthonormal rotation thereof) corresponding to the s non-zero eigenvalues of

$$P = \{p_{ii'}\}, \text{ with } p_{ii'} = \sum_{j=1}^{m} \sum_{k=1}^{n} z_{ijk}z_{i'jk}$$

$\hat{H}$ is the eigenvector matrix (or an orthonormal rotation thereof) corresponding to the t non-zero eigenvalues of

$$Q = \{q_{jj'}\}, \text{ with } q_{jj'} = \sum_{i=1}^{\ell} \sum_{k=1}^{n} z_{ijk}z_{ij'k}$$

$\hat{E}$ is the eigenvector matrix (or an orthonormal rotation thereof) corresponding to the u non-zero eigenvalues of

$$R = \{r_{kk'}\}, \text{ with } r_{kk'} = \sum_{i=1}^{\ell} \sum_{j=1}^{m} z_{ijk}z_{ijk'}$$

$$\hat{C} = \{c_{pqr}\}, \text{ with } \hat{c}_{pqr} = \sum_{i'=1}^{\ell} \sum_{j'=1}^{m} \sum_{k'=1}^{n} \hat{g}_{i'p}\hat{h}_{j'q}\hat{e}_{k'r}z_{i'j'k'}$$

B.2:

On the other hand, if $\hat{G}$, $\hat{H}$, $\hat{E}$, $\hat{C}$ are defined as in B.1, the eigenvalues associated with $\hat{G}$, $\hat{H}$, $\hat{E}$ are different for each matrix separately, and A.3 is satisfied, then

$(\hat{G},\hat{H},\hat{E},\hat{C})$ is the exact unique solution.

▼▼

It should be noted that statement B.2 is not as strong as one would like to have it, as any set of columnwise orthonormal matrices G,H,E, which satisfy A.3 determines an exact solution.

4.4 ALTERNATING LEAST SQUARES ALGORITHM FOR THE TUCKER3 MODEL

Introduction. Obviously we would like to construct an algorithm for the maximization of the fitted sum of squares, *p*, that converges to a global maximum. Unfortunately *p* is the cross-product term of a multivariate polynomial of the sixth degree, and in general it is not possible to prove that methods to solve such nonlinear problems attain a global optimum. In the present case this also seems to be true. We will have to be satisfied with proving that the algorithm outlined below will converge to some stationary point which is not a minimum rather than to a global maximum.

Alternating least squares approach. The method to be described utilizes the so-called alternating least squares (ALS) technique, already referred to above. 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. Further details and references to applications of the ALS approach can, for instance, be found in Young, de Leeuw, & Takane (1980), Young (1981).

In order to see how the ALS approach can be applied in the present context, let us return to the definition of *f*:

$$f\ (G,H,E,C) = \sum_{i=1}^{\ell} \sum_{j=1}^{m} \sum_{k=1}^{n} (z_{ijk} - \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} g_{ip}h_{jq}e_{kr}c_{pqr})^2$$

The sets of parameters are here G, H, E, as C can be derived from the other three (see section 4.3). Minimizing *f* 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 we are in practice maximizing over *p* the problem is still an ALS one.

From the above discussion a rough outline for an algorithm is readily deduced. First choose an arbitrary H_o and E_o and maximize over G to get a new G_1, maximize subsequently over H with the just computed G_1 and E_o fixed to get a new H_1, finally maximize over E with G_1 and H_1 fixed to get a new E_1, and iterate the procedure until - one hopes - convergence. According to *Theorem 4.1* the maximizations are essentially equal to searches for eigenvectors and eigenvalues of matrices of the order ℓ,m, and n respectively. As ℓ,m, and n can be quite large, while s,t, and u are typically very small, say 2,3, or 4, we want to use a technique for solving the eigenvector-eigenvalue problem (or eigenproblem for short) which is particularly efficient in finding only the first few eigenvectors.

A very appropriate technique in this situation is the so-called simultaneous iteration method (or Treppen Iteration) of Bauer-Rutishauser (Rutishauser, 1969).

Thus, the maximization of *p* consists of an, in principle, infinite iteration process, in which at each step three eigenproblems have to be solved. Clearly, solving these eigenproblems by another infinite iteration process has its drawbacks. The whole procedure is likely to become computationally cumbersome. In order to avoid this we perform only one single step towards the solution of the eigenproblems, instead of complete iterations. A similar approach has been applied by De Leeuw and others in a number of cases when using an ALS technique. The experience has been that carrying out the complete iteration to solve the eigenproblem only serves to decrease the overall efficiency of the procedure, while using only one step has no effect on the eventual convergence point (Takane, Young, & De Leeuw, 1977, p.59). They suggest that the reason for this behaviour might be found in the same reasons that often cause relaxation procedures to be more efficient than non-relaxation procedures.

Simultaneous iteration method. Let A be a real n×n symmetric positive definite matrix, and s the desired number of eigenvectors. Furthermore let X be defined as a real n×s matrix which has for its columns the iteration vectors. If we write X after a iterations as X_a, the method of Bauer-Rutishauser is defined as follows.

i. Choose an arbitrary orthonormal X_o.
ii. $Y_a = AX_a$, and
iii. $B_a = Y'_aY_a$.
iv. Solve the eigenproblem for B_i, i.e. determine an orthonormal T_a, and a diagonal matrix L_a with $\ell_1^a \geq \ell_2^a \geq \cdot\cdot \ell_s^a$, such that $T'_aB_aT_a = L_a$, and T_a is the eigenvector matrix of B_a; then define
v. $X_{a+1} = Y_aT_aL_a^{-\frac{1}{2}}$.

Schwartz et al. (1968, p.182-187) show that for a→ ∞, $L_a^{-\frac{1}{2}}$ converges to the matrix with the largest s eigenvalues of A on the diagonal, and the columns of X_a converge to the associated eigenvectors, provided A is positive definite, the columns of X are not orthogonal to one or more of the eigenvectors, and the s-th and (s+1)-th eigenvalues are different. We will write ii. through v. somewhat more concisely by postmultiplying v. with T_a:

$$X_{a+1} = Y_aT_aL_a^{-\frac{1}{2}}T'_a = AX_aB_a^{-\frac{1}{2}} = AX_a(X'_aA^2X_a)^{-\frac{1}{2}}$$

In practical applications we will not perform this postmultiplication as it would yield a rotated version of the principal components rather than the components themselves. For theoretical purposes the postmultiplication is immaterial, but convenient to work with. In Kroonenberg & De Leeuw (1980) the postmultiplication with T'_a for step v. was incorrectly included in the description and implementation of the algorithm presented in that paper. As stated above this does not affect the theoretical results, but in their example the components of what they called the 'unrotated' space were in fact rotated (by some T'_a), and their 'rotated' components were approximately the principal components.

We will define the following function to be used later. When we use in the sequel functions like ϕ we mean to say that

$$\phi(X_a) = X_{a+1} = AX_a(X'_aA^2X_a)^{-\frac{1}{2}}$$

can be computed by carrying out one step of the Bauer-Rutishauser method. It should be noted that the inverse square root of $X'A^2X$ exists, and is uniquely defined, if the expression is positive definite. This implies that in such a case ϕ is well-defined, and it can be proven that ϕ is continuous as well (see the Appendix to Kroonenberg & De Leeuw, 1980).As will be shown below, rather strong convergence theorems can be used for the algorithm to be described if ϕ is continuous. It seems, therefore, worthwhile to take measures in constructing the algorithm to ensure the positive definiteness of $X'A^2X$. An inspection of the method to arrive at X_{a+1} shows that in fact only the inverse square root is taken of the eigenvalues of B_a. One therefore only has to check in each iteration step if all eigenvalues are larger than zero, or in practice larger than some very small number. If one of the eigenvalues is too small, one can restart the iteration procedure with a smaller number of components. There is, however, no guarantee that this will solve the singularity problem. On the other hand, if no singularities have occurred one knows that at each step ϕ must have been uniquely defined and continuous. As we check for the positiveness of the eigenvalues in our programs we will from now on assume that expressions like $X'A^2X$ are positive definite.

In Kroonenberg & De Leeuw (1977) the proofs of the algorithm for the Tucker2 model were formulated without the extra condition that $X'A^2X$ is always positive definite by using so-called point-to-set maps. Although convergence could then be proven, the uniqueness of the solution at each step is no longer assured. The additional complexity of working with point-to-set maps is, however, not really necessary as long as steps are taken to assure the positive definiteness.

TUCKALS3 algorithm. In this subsection we will describe the algorithm to solve the maximization of p. Here Z is again defined as the $\ell \times m \times n$ three-mode data matrix, and s,t, and u will be the desired number of components for the three component matrices. Furthermore the orthonormal matrices G, H, and E will be the matrices whose columns are the iteration vectors. We will write G, H, and E as they are after a iteration steps as G_a, H_a, and E_a. One main iteration step of the TUCKALS3 algorithm is then defined as follows:

(a+1)st step of TUCKALS3

G-substep

$$p^{a}_{ii'} = \sum_{j=1}^{m} \sum_{j'=1}^{m} \sum_{k=1}^{n} \sum_{k'=1}^{n} \sum_{q=1}^{t} \sum_{r=1}^{u} h^{a}_{jq} h^{a}_{j'q} e^{a}_{kr} e^{a}_{k'r} z_{ijk} z_{i'j'k'}$$

$$G_{a+1} = P_a G_a (G'_a P^2_a G_a)^{-\frac{1}{2}}$$

H-sustep

$$q^{a}_{jj'} = \sum_{k=1}^{n} \sum_{k'=1}^{n} \sum_{i=1}^{\ell} \sum_{i'=1}^{\ell} \sum_{r=1}^{u} \sum_{p=1}^{s} e^{a}_{kr} e^{a}_{k'r} g^{a+1}_{ip} g^{a+1}_{i'p} z_{ijk} z_{i'j'k'}$$

$$H_{a+1} = Q_a H_a (H'_a Q^2_a H_a)^{-\frac{1}{2}}$$

E-substep

$$r^{a}_{kk'} = \sum_{i=1}^{\ell} \sum_{i'=1}^{\ell} \sum_{j=1}^{m} \sum_{j'=1}^{m} \sum_{p=1}^{s} \sum_{q=1}^{t} g^{a+1}_{ip} g^{a+1}_{i'p} h^{a+1}_{jq} h^{a+1}_{j'q} z_{ijk} z_{i'j'k'}$$

$$E_{a+1} = R_a E_a (E'_a R^2_a E_a)^{-\frac{1}{2}}$$

As mentioned before, each G, H, and E substep is one step of an inner iteration to find the eigenvectors of P, Q, and R respectively, and together they define one step of the main iteration.

Convergence of the TUCKALS3 algorithm. Before discussing the convergence of the algorithm itself, it is necessary to introduce some new notation.

F: S → S is a function on S; and F defines a complete step of the main iteration; and S is defined as in section 4.3 and *Theorem 4.1*.

$F = F_3 F_2 F_1$ with F_i: S → S for i=1,2,3 such that

$$F_1(G_a,H_a,E_a) = (\phi_1(G_a),H_a,E_a) = (G_{a+1},H_a,E_a)$$

$$F_2(G_{a+1},H_a,E_a) = (G_{a+1},\phi 2(H_a),E_a) = (G_{a+1},H_{a+1},E_a)$$

$$F_3(G_{a+1},H_{a+1},E_{a+1}) = (G_{a+1},H_{a+1},\phi 3(E_a)) = (G_{a+1},H_{a+1},E_{a+1})$$

Thus $F(s_a) = F(G_a,H_a,E_a) = (G_{a+1},H_{a+1},E_{a+1}) = s_{a+1}$

Above we already remarked that ϕ as defined in the previous subsection was a continuous function. Because F is a composite of continuous functions F is continuous as well.

It can be shown that at each step of the main iteration, and at each substep the value of p does not decrease (see the Appendix to Kroonenberg & De Leeuw, 1980). Thus

$$p(F(s_a)) = p(s_{a+1}) \geqq p(s_a)$$

If p is not increased strictly, i.e. $p(F(s_a)) = p(s_a)$, the algorithm stops. In that case (G,H,E) satisfies the necessary conditions of *Theorem 4.3*. Consequently we can assume without loss of generality that the algorithm generates infinite sequences with $p(F(s_a)) > p(s_a)$.

The TUCKALS3 algorithm is a type of algorithm that has been described in the non-linear programming literature, and in that field various theorems about the convergence of algorithms such as ours exist. Appropriate to our case is the following "fixed point" theorem described and proven by d'Esopo (1959):

Theorem 4.3 (d'Esopo's convergence theorem)
Let F, p, S satisfy the following conditions.

1. a. S is a subset of a finite dimensional space,
   b. F is a continuous function,
   c. p is a real function defined and continous for all $s \varepsilon S$,
2. $p(F(s)) \geqq p(s)$,
3. if $p(F(s)) = p(s)$, then $F(s) = s$
4. if the sequence s_o, s_1, ... satisfies $p(s_{a+1}) \geqq p(s_a)$ with $s_a \varepsilon S$, then for every accumulation point $\bar{s}$ of s_o, s_1, ... $F(\bar{s}) = \bar{s}$.

▼▼

In the previous subsections we have discussed all the conditions of *Theorem 4.3*, and we may therefore conclude that it applies to the TUCKALS3 algorithm. As S is a bounded real subspace, any infinite sequence $s_o, s_1, \ldots$ is bounded, and thus the sequences generated by the algorithm are bounded as well. A theorem due to Weierstrass shows that such sequences have at least one accumulation point. It is shown in the Appendix to Kroonenberg & De Leeuw

(1980) that every point $\bar{s}$ such that $F(\tilde{s}) = \bar{s}$ is a stationary point of p, and because we know that at every step p increases, we know that stationary points will not be minima.

Still assuming the positive definiteness of expressions such as $X'A^2X$ we can use a theorem due to Meyer (1976, p.110) to show that $[[s_{a+1} - s_a]] \rightarrow 0$, i.e. the normed difference between the component matrices in successive iterations becomes arbitrary small. It would go too far to present the theorem in detail as it is formulated in terms of point-to-set maps and related concepts (see, however, Kroonenberg & De Leeuw, 1977, p.48 for details). As has been shown by Ostrowski (1966) the set of accumulation points of $\{s_a\}$ consists either of a single point or a continuum. The latter case, however, is very unlikely in practical applications, as is the occurrence of equal eigenvalues in real data matrices. In order to be able to use Meyer's results, it seems sensible to control the convergence of the algorithm both by the objective function, and by the normed difference of all component matrices. In practice it generally turns out that the convergence for the components (i.e. the quantities we especially are interested in) is far slower than the convergence for the objective function.

4.5 NESTING OF COMPONENTS AND INITIALIZATION

Two questions with respect to increasing the number of components need to be answered. First, if we increase the number of components in the modes, does the SS(Fit) always increase as well? Secondly, are the solutions nested, i.e. is the configuration resulting from, say, just two components in the first mode the same as the two-dimensional configuration from the three component configuration of the first mode ?

Nested solutions? The configurations of three-mode principal component analysis in its alternating least squares formulation are in general not nested. This can readily be seen from the algorithm itself. When for instance, the number of components of the first mode is increased, then the functions Q and R in the other substeps are directly influenced by this increased number of components, and therefore different eigenproblems have to be solved. Table 4.1 gives an example of this lack of nesting. On the other hand, the better the fit, the better the nesting, and the more the alternating least squares solution resembles the initialization solution (=Tucker's method solution - see below), the better the nesting.

Table 4.1 *Attachment study: Effect of increasing number of components on percentages explained variation of components*

	components										
solution	episodes			interactive scales			children				sum
	1	2	3	1	2	3	1	2	3	4	
2x2x2	35	23		39	19		49	8			57
2x2x3	37	27		41	22		49	8	6		63
2x2x4	39	27		42	24		48	9	6	2	66
3x3x3	36	27	5	41	25	2	49	10	9		68
3x3x4	37	28	8	45	25	3	49	10	8	6	73

Note: All analyses are based on 4 episodes, 4 interactive scales, and 53 children.

When increasing the number of components for any one mode, the SS(Fit) will in general increase. If the new solution is the global maximum of the SS(Fit) given the number of components, then the new SS(Fit) will never be smaller than the old one. If, however, the iteration stops at a local maximum, this is not necessarily so.

Initialization. In the algorithm we need some G_o, H_o, and E_o to initialize the procedure. It seems sensible to choose them in such a way that they are optimal in some sense. Given that an exact solution exists, an initialization procedure which finds this solution without even entering the main iteration seems a reasonable choice. *Theorem 4.2B.1* tells us that the eigenvectors of P, Q, and R with

$$p_{ii'} = \sum_{j=1}^{m} \sum_{k=1}^{n} z_{ijk}z_{i'jk} \; ; q_{jj'} = \sum_{k=1}^{n} \sum_{i=1}^{\ell} z_{ijk}z_{ij'k}$$

$$r_{kk'} = \sum_{i=1}^{\ell} \sum_{j=1}^{m} z_{ijk}z_{ijk'}$$

should be used for G_o, H_o, and E_o respectively to achieve this purpose. Referring back to section 4.2, the initialization procedure is seen to be identical to Method I proposed by Tucker (1966a, p.297). It is, in general, not necessary to determine these eigenvectors very precisely, because they are only the starting points for the main iteration. Here also the Bauer-Rutishauser procedure is used to compute the eigenvectors.

Upper bounds for SS(Fit). This particular initialization has an additional advantage, because it provides us with upper bounds for the SS(Fit) which is maximized in the main iteration procedure.

Theorem 4.4 (Upper bounds for the SS(Fit)

Let $(\hat{G},\hat{H},\hat{E})$ maximize p as defined above, then

$$p\ (\hat{G},\hat{H},\hat{E}) \leqq \min\ (\sum_p \lambda^{\times}_p, \sum_q \mu^{\times}_q, \sum_r \nu^{\times}_r) \leqq \sum_i \sum_j \sum_k z^2_{ijk}$$

with the $\lambda^{\times}_p$'s, $\mu^{\times}_q$'s, and $\nu^{\times}_r$'s the largest eigenvalues of P, Q, and R respectively, or

$$\text{SS(Fit)} \leqq \min\ \{\text{SS(Fit}_1\text{)},\text{SS(Fit}_2\text{)},\text{SS(Fit}_3\text{)} \leqq \text{SS(Total)}$$

The proof of this theorem can be found in the Appendix 4.2. In words the theorem says that the fitted sum of squares of the main iteration procedure, (SS(Fit), can never be larger than the smallest of the fitted sum of squares resulting from standard principal component analyses on the cross-product matrices from each of the modes SS(Fit_m), m=1,2,3. These SS(Fit_m)s are in turn always smaller than or equal to the total sum of squares, SS(Total).

The practical importance of this result is that it is possible to gauge how well the iteration procedure has succeeded in finding an optimal solution, given the number of components for each mode. It can also point the way to improving the overall fit for the data, especially in the case where the choice of number of components for one mode was such that the fit possible with that number of components is very much smaller than the fit for the other modes. Suppose, for example, that the relative fit for the first mode is only 0.50, while for the other two modes a relative fit is possible of 0.80. The overall relative fit will then be necessarily no greater than 0.50 according to *Theorem 4.4*. The only really effective way to increase the overall relative fit is to include more components for the first mode, as it provides the smallest upper bound.

4.6 ALTERNATING LEAST SQUARES ALGORITHM FOR TUCKER2 MODEL

After the extensive discussion of the solution for the Tucker3 model, we can be very brief about the solution for the Tucker2 model. Its estimation poses no new problems; all one has to do is delete the E substep from the TUCKALS3 algorithm, and insert the k×k identity matrix for E in the other substeps. With these provisions the TUCKALS3 algorithm can be used for the Tucker2 model. Computationally it is, however, more efficient to solve the model by an analogue of the TUCKALS3 algorithm. The proofs for the TUCKALS2 algorithm are entirely parallel to those for the TUCKALS3 ones, and therefore need not be discussed again. Explicit proofs have been given in Kroonenberg & De Leeuw (1977) using a more general formulation of convergence, as was already mentioned above.

TUCKALS2 algorithm. We will only present the TUCKALS2 algorithm, and we will not discuss it in any detail, nor will we do so with the model. We will return to the model in Chapter 5 when we discuss the problem of transformations of the extended core matrix, and the uses of such transformations. In later chapters we will present a number of examples of the Tucker2 model.

(a+1)-th step of TUCKALS2

G-substep: $p^a_{ii'} = \sum_{k=1}^{n} \sum_{j=1}^{m} \sum_{j'=1}^{m} \sum_{q=1}^{t} h^a_{jq} h^a_{j'q} z_{ijk} z_{i'j'k}$

$$G_{a+1} = P_a G_a (G'_a P^2_a G_a)^{-\frac{1}{2}}$$

$$\text{H-substep: } q^{a}_{jj'} = \sum_{k=1}^{n} \sum_{i=1}^{\ell} \sum_{i'=1}^{\ell} \sum_{p=1}^{s} g^{a+1}_{ip} g^{a+1}_{i'p} z_{ijk} z_{i'j'k}$$

$$H_{a+1} = Q_a H_a (H'_a Q^2_a H_a)^{-\frac{1}{2}}$$

4.7 COMPUTATIONAL ACCURACY AND PROPAGATION OF ERRORS

Developing formal expressions for the numerical accuracy of the programs is somewhat difficult considering the complexity of the algorithm. Especially modelling the propagation of rounding errors in successive iterations is not straightforward. The orthonormalization of the components at each iteration substep assures, however, that the errors of the components never run out of hand. In addition, the manipulation of very large or very small numbers is avoided by rescaling the overall variation of the data set to be analysed to $\ell \times m \times n$ (= volume of the data cube). Such scaling does not affect the components themselves, but only the absolute sizes of the eigenvalues and the elements of the core matrix (see also section 6.6). However, those quantities are in generally best interpreted as percentages of the total variation (see section 6.9), and in that respect the rescaling does not affect interpretation. The rescaling has the additional advantage that the convergence criteria are more or less equally strict for all problems.

In the examples below we will give some indication of the numerical accuracy of the TUCKALS3 program, and of the effect of introducing error in data with a known structure. No attempt has been made to treat these problems analytically or exhaustively.

Hilbert cubes and replicated Hilbert matrices. To our knowledge no three-way matrices with known eigenvalues have been published; therefore we have taken some substitutes, i.e. the Hilbert cube and the replicated Hilbert matrix.
The Hilbert matrix of the order n, H_n, is defined as $H_n = \{h^n_{ij}\}$ with $h^n_{ij} = 1/(i+j-1)$ i,j=1,...,n. Analoguously the *three-mode Hilbert cube* can be defined as $H_n = \{h^n_{ijk}\}$ with $h^n_{ijk} = 1/(1+j+k-2)$

i,j,k=1,...,n. This cube is three-way symmetric (i.e. symmetric around the body diagonal). The component matrices derived from it should, therefore, be identical, and the core matrix should be three-way symmetric as well, provided the same number of components is taken for every mode. The identities and symmetry were realized until at least the tenth decimal. Several tests have been carried out with the initialization procedure described in section 4.5, and with random start matrices. In both cases the same results were obtained, be it that a solution was found almost without using the main iteration procedure when the initialization routine was employed. In other words, as was pointed out in section 4.5, if there exists an exact solution, the initialization procedure will find it first.

These results are, of course, no guarantee that the solutions are correct, they only have the desired form. But lacking analytical results it is the best we can do for the moment. For future reference the results for the Hilbert cube of order 4 are given in Table 4.2.

Table 4.2 *Hilbert cube of order 4 (with initialization procedure)*

component matrices and standardized weights

	1	2	3
1.	0.7383 103	0.6429 093	-0.2009 515
2.	0.4805 923	-0.3000 370	0.7458 193
3.	0.3666 971	-0.4813 999	-0.0569 134
4.	0.2991 024	-0.5146 839	-0.6325 606
stand. weight	0.9819 994	0.0179 479	0.0000 607

standardized core matrix

c(1,1,1) = 0.9866 264
c(2,1,1) = -0.0022 338
c(3,1,1) = 0.0005 113
c(2,2,1) = 0.0923 996
c(3,2,1) = 0.0008 774
c(3,3,1) = 0.0360 099
c(2,2,2) = 0.0283 716
c(3,2,2) = -0.0054 683
c(3,3,2) = 0.0010 898
c(3,3,3) = -0.0023 065

To have at least a partial check on the accuracy and the correctness of the algorithm and the program, a *replicated Hilbert matrix* of order 4 was used, i.e. $Z = \{z_{ijk}\}$ with $z_{ijk} = h^n_{ijk} = 1/(i+j-1)$ i,j,k=1,...,n. Each frontal plane is thus equal to the same Hilbert matrix, and there are n of them. It is simple to show that $G = H = K_n$, with K_n the eigenvector matrix of the Hilbert matrix of order n, in which the eigenvectors are standardized at 1. E will be one dimensional with elements equal to $(1/n)^{\frac{1}{2}}$. In Table 4.3 the results for the replicated Hilbert matrix of order 4 are compared with the results given in Gregory & Karney (1969).

Table 4.3 *Replicated Hilbert matrix of order 4*

eigenvectors for mode 1 and 2

	1			2			3		
	stem	G&K	T3	stem	G&K	T3	stem	G&K	T3
1.	0.79260	829	832	0.58207	570	578	-0.17918	629	597
2.	0.45192	312	306	-0.37050	218	262	0.74191	779	812
3.	0.32241	639	641	-0.50957	863	853	-0.10022	814	959
4.	0.25216	117	119	-0.51404	827	797	-0.63828	253	200
stand. weight	0.98742	851	902	0.01255	157	159	0.00001	992	992

G&K = Gregory & Karney (1969); T3 = TUCKALS3; stem = same for both

With the replicated Hilbert matrix, as with the Hilbert cube, it was not possible to find the smallest root (1.87×10^{-8}). In fact this eigenvalue was so small that the 4x4 matrices P_a and Q_a (see section 4.4), of which the eigenvectors had to be computed, were considered singular by the program. The restart procedure, which reduces the number of components by the number of eigenvalues considered to be too small (see section 4.4), took care of this situation, and caused the program to start again with the reduced number of components.

Propagation of errors in similarity judgements. A small Monte Carlo experiment was conducted to gain some insight in the error propagation in the TUCKALS3 method. The data from the example presented in detail in Chapter 2 were subjected to various degrees of perturbation. More in particular, the data (unlike the real situation) were considered to be judgements on a nine-point scale of the similarity between the row and column points, with the row points serving as standards. To introduce error the observed similarities were taken as the means of symmetric discrete distributions with varying heaviness of the tails. Thus if a stimulus combination (i,j) was scored, for instance $z_{ij} = 7$, then in error condition *e1* a score is generated from the discrete distribution:

$\Pr(z_{ij} = 6) = \Pr(z_{ij} = 8) = .05; \Pr(z_{ij} = 7) = .90;$
$\Pr(z_{ij} = 1 \text{ or } 2 \text{ or } 3 \text{ or } 4 \text{ or } 5 \text{ or } 9) = 0.$

Table 4.4 shows how the end points were dealt with.

Table 4.4 *Error structures for the De Gruijter data*

		generated score, Zg							
		1	2	3	4	...	7	8	9
	1	95	5	0	0	...	0	0	0
original	2	5	90	5	0	...	0	0	0
score	3	0	5	90	5	...	0	0	0
Z_o	:	:	:	:	:	:::	:	:	:
	8	0	0	0	0	...	5	90	5
	9	0	0	0	0	...	0	5	95

condition *e1*

	generated score, Z_g							
	1	2	3	4	5	6	7	
1	75	10	10	5	0	0	0	...
2	25	50	10	10	5	0	0	...
3	15	10	50	10	10	5	0	...
4	5	10	10	50	10	10	5	...
5	0	5	10	10	50	10	10	...
:	:	:	:	:	:	:	:	...

condition *e5*

$$p_{ij} = \Pr\,(Z_g = j \mid Z_o = i)$$

The underlying idea in using this error structure was that a subject generally produces the number he intends (his 'true' score), but there is a non-zero probability that the similarity probably was meant to be higher or lower, and large differences from the intended score are less likely than smaller ones. Increasing errors, i.e. distributions with longer tails, imply increasing vagueness of the judgments. The results of the effects of increasing error in the data is summarized in Table 4.5, and Figure 4.1.

Table 4.5 *Effects of perturbations on the De Gruijter data*

Absolute differences with the unperturbed solution

components for mode 1

1					2					3				
e0	*e1*	*e3*	*e5*	*e8*	*e0*	*e1*	*e3*	*e5*	*e8*	*e0*	*e1*	*e3*	*e5*	*e8*
48	0	3	1	2	-25	0	5	3	5	- 4	0	5	4	13
48	0	4	3	8	-17	0	2	4	14	4	2	1	5	8
43	0	1	2	8	18	3	2	4	2	15	4	1	8	5
1	0	2	0	11	50	1	0	4	11	- 9	1	7	5	12
-14	0	3	2	9	30	2	1	3	4	40	0	0	1	4
-20	2	2	0	0	22	7	3	4	8	-62	6	5	4	5
-22	0	1	2	4	29	1	1	11	4	1	1	10	5	11
-33	1	3	2	1	-33	0	4	2	14	29	3	1	6	19
-33	1	2	3	8	-28	3	4	6	3	34	0	7	3	11
-17	1	1	2	4	-46	3	3	4	3	-47	8	6	5	2
MAD	0.5	2.2	1.8	5.5		2.0	2.5	4.5	6.9		2.5	4.3	4.6	9.0
% 61	56	52	45	28	21	20	20	17	14	11	11	11	9	7

components for mode 3

1									
e0	*e1*	*e3*	*e5*	*e8*	*e0*	*e1*	*e3*	*e5*	*e8*
42	0	2	3	4	68	6	9	44	67
42	1	2	0	2	34	1	3	11	42
40	2	2	0	3	-59	6	17	27	106
38	1	3	0	1	-19	8	13	5	24
42	0	3	0	2	-10	10	51	16	70
39	0	0	3	3	-18	12	16	5	56
MAD	0.7	2.0	1.0	2.5		7	18	18	68
% 91	86	81	70	46	1	1	1	1	3

% variation accounted for SS(Fit)/SS(Total)	
e0	98
e1	87
e3	83
e5	71
e8	49

Notes:

e0: 1.00
e1: .05 .90 .05
e3: .05 .10 .70 .10 .05
e5: .05 .10 .10 .50 .10 .10 .05
e8: .10 .10 .10 .10 .20 .10 .10 .10 .10

MAD: mean absolute difference
% : percentage variation accounted for

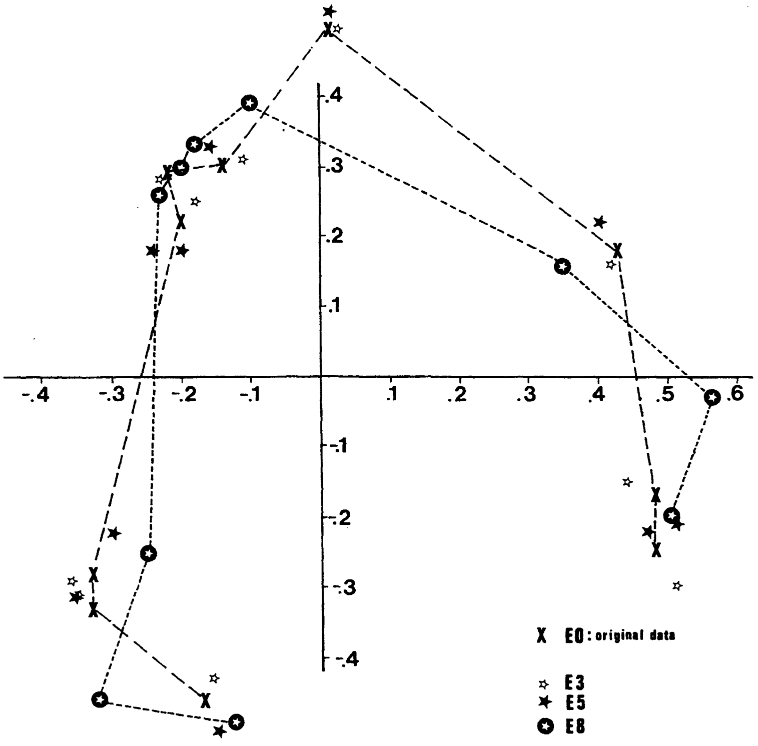


Fig. 4.1 *Party similarity study: Perturbed first mode component loadings*

The overall spatial organization of the points in the first (and second) mode is not fundamentally altered by the errors introduced, especially when the weights or percentages of variation accounted for by the components is sizeable. When they are not, the distortion can be very serious as the second component of the third mode shows. It is not unlikely that the data set and error structures chosen make a far better recovery possible than generally will be the case. The structure in the data is extremely well accounted for by the solution. The error structures are very regular, and favour good recovery, but unlike in some other studies they bear some relation to reality.

4.8 CONCLUSION

In this chapter we have shown that it is possible to solve the estimation problem of the Tucker3 and Tucker2 model both exactly and approximately by alternating least squares methods. The essential difference between the ALS methods presented in this chapter and Tucker's methods, is that the ALS procedures take into account the reduction over the other modes, while Tucker's methods do not. Furthermore, the use of least squares loss functions allows assessment of the model, analysis of residuals, and provides a number of attractive interpretational possibilities as we shall see in later chapters. The examples show that the method as programmed is accurate, and, for data with a well-defined structure, is robust against errors which could have arisen during the production of the data.

APPENDIX 4.1 PROOF THAT SS(TOT) = SS(FIT) + SS(RES)

Let Z be a three-mode data matrix, and G, H, and E arbitrary but fixed columnwise orthonormal matrices (see section 1.5). Let $\hat{C}$, a three-mode core matrix, minimize the loss function

$$[[Z - \tilde{Z}]]^2 = \sum_i \sum_j \sum_k (z_{ijk} - \tilde{z}_{ijk})^2 \text{ with} \qquad \text{(A4.1)}$$

$$\tilde{z}_{ijk} = \sum_p \sum_q \sum_r g_{ip} h_{jq} e_{kr} c_{pqr},$$

then $\hat{C} = \{\hat{c}_{pqr}\}$ is equal to (see section 4.3)

$$\hat{c}_{pqr} = \sum_i \sum_j \sum_k g_{ip} h_{jq} e_{kr} z_{ijk}.$$

The $\hat{Z}$ which minimizes the loss function (A4.1) is

$$\hat{z}_{ijk} = \sum_p \sum_q \sum_r g_{ip} h_{jq} e_{kr} \hat{c}_{pqr} = \sum_p \{\sum_q \sum_r g_{ip} h_{jq} e_{kr} \hat{c}_{pqr}\} = \sum_p t_{ijkp}$$

In section 4.3 it was shown that $\hat{c}_{pqr}$ is uniquely defined, thus for any $c^*_{pqr} = \lambda_p \hat{c}_{pqr}$, (A4.1) attains it minimum only if $\lambda_p = 1$ for all p. Let z^*_{ijk} be defined as

$$z^*_{ijk} = \sum_p \sum_q \sum_r g_{ip} h_{jq} e_{kr} c^*_{pqr} = \sum_p \sum_q \sum_r \lambda_p g_{ip} h_{jq} e_{kr} c_{pqr} =$$

$$= \sum_p \lambda_p (\sum_q \sum_r g_{ip} h_{jq} e_{kr} c_{pqr}) = \sum_p \lambda_p t_{ijkp}.$$

The loss function (A4.1) becomes with z^*_{ijk}

$$\sum_i \sum_j \sum_k (z_{ijk} - z^*_{ijk})^2 =$$

$$\sum_i \sum_j \sum_k z^2_{ijk} - 2 \sum_i \sum_j \sum_k \{z_{ijk} \cdot \sum_p \lambda_p t_{ijkp}\} + \sum_i \sum_j \sum_k (\sum_p \lambda_p t_{ijkp})^2 \qquad \text{(A4.2)}$$

(A4.2) has a minimum if all $\lambda_p = 1$ simultaneously, thus

$$\frac{\delta}{\delta\lambda_p}\{\sum_i \sum_j \sum_k (z_{ijk} - z_{ijk})^2\} = 0 \text{ for all } \lambda_p = 1 \rightarrow$$

$$- \sum_i \sum_j \sum_k z_{ijk} t_{ijkp} + \sum_i \sum_j \sum_k (\sum_{p'} \lambda_{p'} t_{ijkp'}) t_{ijkp} = 0 \text{ for } \lambda_p = 1$$

$$- \sum_i \sum_j \sum_k z_{ijk} t_{ijkp} + \sum_i \sum_j \sum_k (\sum_{p'} t_{ijkp'}) t_{ijkp} = 0$$

$$\sum_i \sum_j \sum_k z_{ijk} t_{ijkp} = \sum_i \sum_j \sum_k [t_{ijkp} \{\sum_{p'} t_{ijkp'}\}] \text{ for all p simultaneously}$$

$$\sum_i \sum_j \sum_k \{\sum_p z_{ijk} t_{ijkp}\} = \sum_i \sum_j \sum_k [\sum_p \{t_{ijkp} (\sum_{p'} t_{ijkp'})\}]$$

$$\sum_i \sum_j \sum_k z_{ijk} (\sum_p t_{ijkp}) = (\sum_p t_{ijkp})^2$$

$$\sum_i \sum_j \sum_k z_{ijk} \hat{z}_{ijk} = \sum_i \sum_j \sum_k \hat{z}^2_{ijk}.$$

Thus using this last result, we may conclude that

$$\sum_i \sum_j \sum_k (z_{ijk} - \hat{z}_{ijk})^2 = \sum_i \sum_j \sum_k z^2_{ijk} - 2\sum_i \sum_j \sum_k z_{ijk} \hat{z}_{ijk} + \sum_i \sum_j \sum_k \hat{z}^2_{ijk}$$

$$= \sum_i \sum_j \sum_k z^2_{ijk} - \sum_i \sum_j \sum_k \hat{z}^2_{ijk}$$

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

or SS(Tot) = SS(Fit) + SS(Res).

▼▼

APPENDIX 4.2 BOUNDS FOR THE SS(FIT)

Theorem 4.4

Let G, H, and E maximize *p* as defined in *Theorem 1*, then

$$p(G,H,E) \le \min\ (\sum_p \lambda_p, \sum_q \mu_q, \sum_r \nu_r) \le \sum_i \sum_j \sum_k z^2_{ijk}$$

with the λ_p, μ_q, and ν_r the largest eigenvalues of *P*, *Q*, and *R* (see section 4.2 for definitions), respectively.

Proof:

Define $Z \varepsilon R^{\ell \times mn}$, and $Z_i \varepsilon R^{m \times n}$, $i=1,\dots,\ell$.

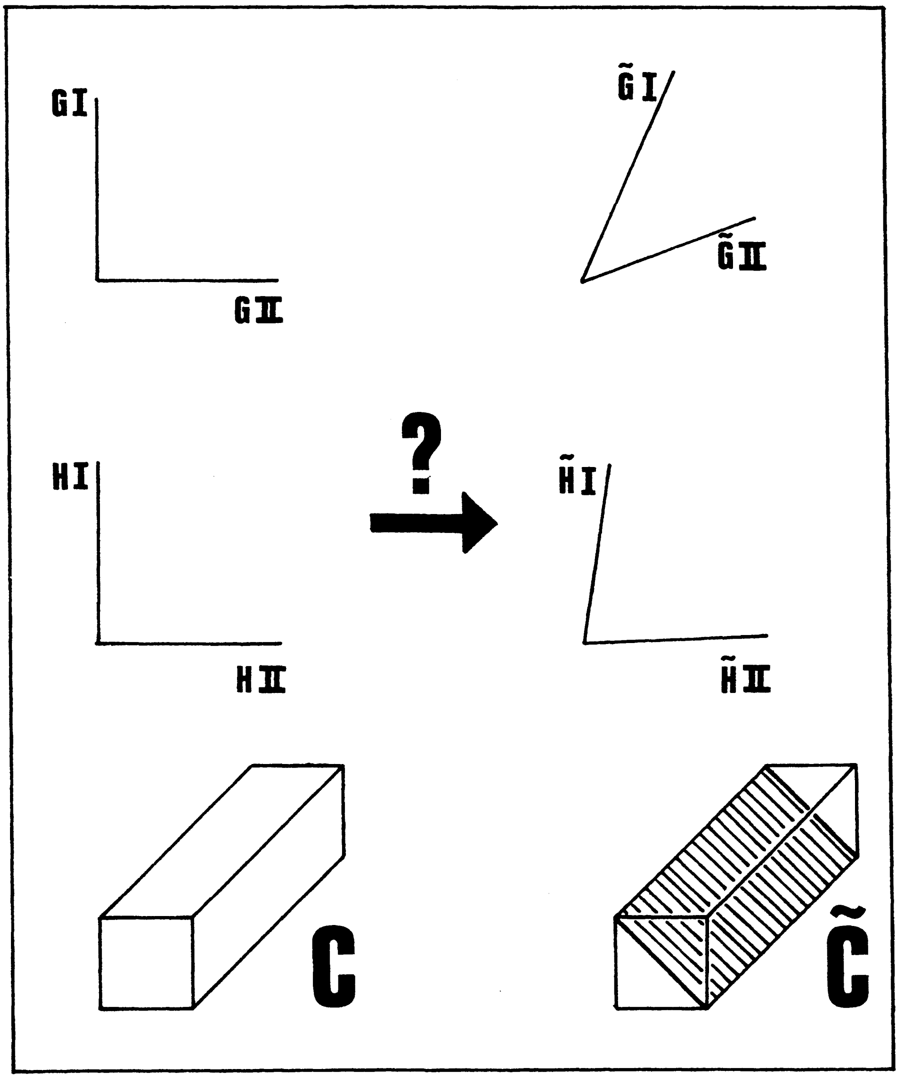
$$\begin{aligned} p(G,H,E) &= \text{tr } G'\{Z(HH' \otimes EE')Z'\}G \\ &\le \text{tr } Z(HH' \otimes EE')Z' = \text{tr } \sum_i H'Z_iEE'Z_i'H = \text{tr } H'\{\sum_i Z_iEE'Z_i'\}H \\ &\le \text{tr } \sum_i Z_iEE'Z_i' = \text{tr } E\{\sum_i Z_iZ_i'\}E = \text{tr } E'RE \\ &\le \sum_{r=1}^{u} \nu_r, \text{ with } \nu_r \text{ the u largest eigenvalues of } R \\ &\le \sum_{r=1}^{n} \nu_r = \sum_i \sum_j \sum_k z^2_{ijk}. \end{aligned}$$

The inequalities follow from the so-called separation theorem, which states that the eigenvalues of an arbitrary section, say E'*R*E, of a symmetric matrix *R* separate the eigenvalues of *R* itself (cf. Householder , 1964).

The analoguous results hold clearly for the other two formulations of *p*(G,H,E), and combining these results establishes the theorem. ▼▼

TRANSFORMATIONS OF CORE MATRICES

5



5.1 INTRODUCTION

In this chapter we will discuss the problem of diagonality of the extended core matrix of the Tucker2 model. As the subject is still being explored a definitive treatment cannot yet be given. In section 3.2 we indicated that the orthonormal restrictions put on the Tucker2 and Tucker3 model were necessary to identify the equations for solving the minimization problem, but that these restrictions could be made without loss of generality. In fact, we use the transformational freedom to choose such component matrices that a convenient and efficient algorithm could be devised. Once a solution is found we can drop the restrictions, and use any (non-singular) transformation we like on the components or core matrix, provided the appropriate counter-rotations are performed as well.

One way to use the transformational freedom in three-mode principal component models is to search for transformations that create "simple structures" in the component matrices. Another approach is to search for transformations that induce a simple structure into the core matrix. In this chapter we will look at transformation procedures which attempt the latter by searching for those transformations which diagonalize the core matrix. By doing this we will restrict ourselves to situations where the number of components is equal for the first and second mode (i.e. s=t). This is in itself no restriction as the procedures to be described are such that s and t will become equal in any event.

The procedures outlined in this chapter can be applied to the frontal core planes from both Tucker2 (T2) and Tucker3 (T3) models. In section 3.4 we used the idea of transforming the T3 core matrix to diagonality to show that in specific cases the Tucker3 model has the same form as the PARAFAC1/CANDECOMP model. In this chapter we will not treat these and other issues connected with the Tucker3 model, but leave them for another study. We will concentrate on the diagonality of the extended core matrix of the Tucker2 model.

We will discuss two transformation procedures to optimize diagonality: one using orthonormal transformations, and the other using non-singular transformations. One of the purposes of the transformations to diagonality is to investigate whether the data fit more restricted models such as INDSCAL, which *require* the extended core matrix to be diagonal. We will, however, not present detailed comparisons of solutions from TUCKALS2 plus transformations, and solutions from INDSCAL, as for instance MacCallum (1976a,b) did. In section 11.2 such comparisons are made for data from the *Cola study* (Schiffman, Reynolds, & Young, 1981), but the example is not very demanding.

Another related aim of the transformations is to simplify the interpretation, as a diagonal core matrix with its zero off-diagonal elements displays a typical 'simple structure'. In the case of non-singular transformations, this simplicity is bought by the non-orthogonality of the components in the first and second mode. At present, it is not clear which of the two properties is more desirable in specific situations, the more so because especially the non-singular transformation still poses a number of unsolved interpretational problems.

An extended three-mode core matrix $C = (C_1,\dots,C_n)^*$ is defined to be *diagonal* if for each k (k=1,..,n) $c_{pqk} = 0$ for all $p \neq q$. Note that we do not require c_{ppk} to be unequal to zero, but not all c_{ppk} for all k and fixed p may be zero at the same time, because then we would have ended up with one component less in the first and second mode. From now on we will assume that for each p there is always a k, for which c_{ppk} is not zero.

The procedures outlined in section 5.2 and 5.3, and compared in section 5.4, will be applied to two examples in section 5.5.

The discussion of the theory and application of the transformations will be rather incomplete, primarily because the experience with these methods is still very limited.

---

[*] For convenience, we will write in this chapter C instead of $\tilde{C}$ for the extended core matrix.

5.2 ORTHONORMAL TRANSFORMATIONS

Problem and solution. In this section we will outline a procedure to transform an extended core matrix to optimal diagonality - in a least squares sense - by using two orthonormal transformations. The procedure was first presented in Kroonenberg & De Leeuw (1977, Appendix One).

Diagonality problem (ON)

Let $C = (C_1,C_2,\ldots,C_n)$ with $C_k \varepsilon R^{s\times s}$ $(k=1,\ldots,n)$ be given. Find the (s×s) orthonormal matrices K and L and $D = (D_1,D_2,\ldots,D_n)$ with D_k diagonal $(k=1,\ldots,n)$, such that

$$\sigma(K,L,D) = \sum_{k=1}^{n} \text{tr } (D_k - KC_kL')'(D_k - KC_kL') \tag{5.1}$$

is as small as possible.

Theorem 5.1

Let $C = (C_1,C_2,\ldots,C_n)$ with $C_k \varepsilon R^{s\times s}$ $(k=1,\ldots,n)$, and the problem ON be given. Then

$$\hat{K} = \hat{U}(\hat{U}'\hat{U})^{-\frac{1}{2}} \text{ with } \hat{U} = \sum_{k=1}^{n} \hat{D}_k\hat{L}C_k',$$

$$\hat{L} = \hat{V}(\hat{V}'\hat{V})^{-\frac{1}{2}} \text{ with } \hat{V} = \sum_{k=1}^{n} \hat{D}_k\hat{K}C_k, \text{ and}$$

$$\hat{D}_k = \text{diag } (\hat{K}C_k\hat{L}') \; (k=1,\ldots,n)$$

solve the diagonality problem ON.

Proof:

The solution of problem ON is equivalent to the minimization of

$$\tilde{\sigma}(K,L,D,M,N) = \sum_{k=1}^{n} \text{tr } (D_k - KC_kL')'(D_k - KC_kL') -$$

$$-\tfrac{1}{2} \text{ tr } M(K'K - I_s) - \tfrac{1}{2} \text{ tr } N(L'L - I_s) \tag{5.2}$$

where M and N are symmetric matrices of Lagrange multipliers. Let σ be rewritten in the two following ways

$$1.\ \sigma(K,L,D) = \sum_{k=1}^{n} \operatorname{tr} D_k'D_k - 2\sum_{k=1}^{n} \operatorname{tr} K'D_kLC_k' + \sum_{k=1}^{n} \operatorname{tr} C_k'C_k$$

$$= \sum_{k=1}^{n} \operatorname{tr} D_k'D_k - 2 \operatorname{tr} K' \sum_{k=1}^{n} D_kLC_k' + \sum_{k=1}^{n} \operatorname{tr} C_k'C_k$$

$$= \sum_{k=1}^{n} \operatorname{tr} D_k'D_k - 2 \operatorname{tr} K'U + \sum_{k=1}^{n} \operatorname{tr} C_k'C_k$$

$$2.\ \sigma(K,L,D) = \sum_{k=1}^{n} \operatorname{tr} D_k'D_k - 2 \operatorname{tr} L' \sum_{k=1}^{n} D_kKC_k + \sum_{k=1}^{n} \operatorname{tr} C_k'C_k$$

$$= \sum_{k=1}^{n} \operatorname{tr} D_k'D_k - 2 \operatorname{tr} L'V + \sum_{k=1}^{n} \operatorname{tr} C_k'C_k.$$

Substituting these expressions succesively into (5.2) and differentiating $\tilde{\sigma}$ with respect to K, M, L, N, D, and equating these partial derivatives to zero, we obtain the following set of equation from the stationary equations:

$$\hat{U} = \hat{K}\hat{M} \quad \text{and } K'K = I_s \tag{5.3}$$

$$\hat{K}'\hat{K} = I_s \tag{5.4}$$

$$\hat{V} = \hat{L}\hat{N} \tag{5.5}$$

$$\hat{L}'\hat{L} = I_s \tag{5.6}$$

$$\hat{D}_k = \operatorname{diag}\ (\hat{K}C_k\hat{L}') \quad (k=1,\ldots,n) \tag{5.7}$$

Premultiplying (5.3) with its transpose, and using (5.4) we get

$$\hat{U}'\hat{U} = \hat{M}\hat{K}'\hat{K}\hat{M} = \hat{M}'\hat{M} = \hat{M}^2 \;=\; \hat{M} = (\hat{U}'\hat{U})^{\frac{1}{2}} \tag{5.8}$$

Substituting (5.8) into (5.3), and isolating $\hat{K}$:

$$\hat{K} = \hat{U}(\hat{U}'\hat{U})^{-\frac{1}{2}}\ .$$

Analoguously

$$\hat{L} = \hat{V}(\hat{V}'\hat{V})^{-\frac{1}{2}}.$$

Substituting these $\hat{K}$ and $\hat{L}$ into (5.7) is sufficient to find $\hat{D}_k$ for $k=1,\ldots,n$. ▼▼

Algorithm. From the above theorem a computational procedure can easily be derived, resulting in an alternating least squares algorithm similar to the TUCKALS2 algorithm. A main iteration step of the ON-algorithm will be defined as:

K substep

$$U_a = \sum_{k=1}^{n} D_k^{(a)} L_a C_k'$$

$$K_{a+1} = U_a (U_a'U_a)^{-\frac{1}{2}}$$

L substep

$$V_a = \sum_{k=1}^{n} D_k^{(a)} K_{a+1} C_k$$

$$L_{a+1} = V_a (V_a'V_a)^{-\frac{1}{2}}$$

D substep

$$D_k^{(a+1)} = \text{diag}\left[K_{a+1} C_k L_{a+1}'\right] \quad (k=1,\dots,n)$$

Both $(U_a'U_a)^{-\frac{1}{2}}$ and $(V_a'V_a)^{-\frac{1}{2}}$ can be computed in the same manner as in the equivalent expression in the TUCKALS2 algorithm, i.e. by solving the eigenproblem of $U_a'U_a$ and $V_a'V_a$, and taking the inverse square root of the eigenvalues. Problems of non-uniqueness occur here too, in the case of singularities in $U_a'U_a$ and $V_a'V_a$, but these can be overcome in the same manner as in the TUCKALS2 algorithm.

For proof of the convergence one can adapt the proof for the TUCKALS algorithms given in section 4.4 and in Kroonenberg & De Leeuw (1980).

5.3 NON-SINGULAR TRANSFORMATIONS

Problem and solution. The procedure presented in this section is, in fact, nothing but the CANDECOMP procedure, as outlined in Carroll & Chang (1970). As set forward in section 3.3 the CANDECOMP model

$$z_{ijk}=\sum_{p=1}^{n} g_{ip}h_{jp}c_{ppk}$$

is used to find the best approximate decomposition for certain scalar-product data. CANDECOMP can be used for non-singular transformation of an extended core matrix with different first and second mode; a similar proposal has been made by Cohen (1974, 1975) to use INDSCAL on the extended core matrix from three-mode scaling.

The problem of achieving optimal diagonality of the extended core matric using non-singular transformations can be formulated as the

Diagonality problem (NS)

Let $C = (C_1, C_2, \ldots, C_n)$ with $C_k \in R^{s\times s}$ $(k=1,\ldots,n)$ be given. Find the non-singular $A \in R^{s\times s}$ and $B \in R^{s\times s}$, and $D = (D_1, D_2, \ldots, D_n)$ with D_k is diagonal $(k=1,\ldots,n)$, such that

$$\tau(A,B,D) = \sum_{k=1}^{n} \text{tr } (C_k - AD_kB')'(C_k - AD_kB') \qquad (5.9)$$

is as small as possible.

Theorem 5.2

Let $C_1 = (C_1, C_2, \ldots, C_n)$ with $C_k \in R^{s\times s}$ $(k=1,\ldots,n)$, and the diagonality problem NS be given. Then

$$\hat{A} = \hat{U}_A'\hat{V}_A^{-1} \text{ with } \hat{U}_A = \sum_{k=1}^{n} \hat{D}_k\hat{B}'C_k' \text{ and } \hat{V}_A = \sum_{k=1}^{n} \hat{D}_k\hat{B}'\hat{B}\hat{D}_k,$$

$$\hat{B} = \hat{U}_B'V_B^{-1} \text{ with } \hat{U}_B = \sum_{k=1}^{n} \hat{D}_k\hat{A}'C_k \text{ and } \hat{V}_B = \sum_{k=1}^{n} \hat{D}_k\hat{A}'\hat{A}\hat{D}_k, \text{ and}$$

$$\hat{D}_k = (\hat{d}_1^k, \ldots, \hat{d}_s^k) \text{ with } \hat{d}_p^k = \sum_{q=1}^{s} (\hat{B}'\hat{C}_k'A)_{qq} \quad (\hat{A}'\hat{A}x\hat{B}'\hat{B})_{pq}^{-1},$$

where "x" is the element-wise product of $(p=1,\ldots,s\,;$ $k=1,\ldots,n)$ two matrices, solve the diagonality problem NS, and the minimum is equal to

$$\tau(\hat{A},\hat{B},\hat{D}) = \sum_{k=1}^{n} \text{tr } C_k'C_k - \text{tr } \hat{B}\hat{U}_B' = \sum_{k=1}^{n} \text{tr } C_k'C_k - \text{tr } \hat{A}\hat{U}_A'.$$

Proof:

$\tau(A,B,D) = \sum_{k=1}^{n} \text{tr } (C_k - AD_kB')'(C_k - AD_kB')$ can be written in two ways:

$$1.\ \tau(A,B,D) = \sum_{k=1}^{n} \text{tr } C_k'C_k - 2\sum_{k=1}^{n} \text{tr } BD_kA'C_k + \sum_{k=1}^{n} \text{tr } BD_kA'AD_kB'$$

$$= \sum_{k=1}^{n} \text{tr } C_k'C_k - 2 \text{ tr } B\{\sum_{k=1}^{n} D_kA'C_k\} + \text{tr } B\{\sum_{k=1}^{n} D_kA'AD_k\}B'$$

$$= \sum_{k=1}^{n} \text{tr } C_k'C_k - 2\text{tr } BU_B + \text{tr } BV_BB' \qquad (5.10)$$

with $U_B = \sum_{k=1}^{n} D_kA'C_k$, and

$V_B = \sum_{k=1}^{n} D_kA'AD_k$

$$2.\ \tau(A,B,D) = \sum_{k=1}^{n} \text{tr } C_k'C_k - 2 \text{ tr } AU_A + \text{tr } AV_AA' \qquad (5.11)$$

with $U_A = \sum_{k=1}^{n} D_kB'C_k'$, and

$V_A = \sum_{k=1}^{n} D_kB'BD_k$

Differentiating τ with respect to A, B, and D leads to

$$\frac{\delta}{\delta A} \tau(A,B,D)\Big|_{\hat{A}} = -2U_A' + 2AV_A\Big|_{\hat{A}} = 0$$

$$\frac{\delta}{\delta B} \tau(A,B,D)\Big|_{\hat{B}} = -2U_B' + 2BV_B\Big|_{\hat{B}} = 0$$

$$\frac{\delta}{\delta D_k} \tau(A,B,D)\Big|_{\hat{D}_k} = -2\text{diag } (B'C_k'A) + 2 \text{ diag } (A'AD_kB'B)\Big|_{\hat{D}_k} = 0$$

$$(k=1,\ldots,n)$$

which gives as solution of the stationary equations

$$\hat{A} = \hat{U}_A' \hat{V}_A^{-1} \qquad (5.12)$$

$$\hat{B} = \hat{U}_B' \hat{V}_B^{-1}, \text{ and} \qquad (5.13)$$

$$d_p^k = \sum_{q=1}^{s} (B'C_k'A)_{qq} (A'A \text{ x } B'B)_{pq}^{-1} \quad (p=1,\ldots,s;k=1,\ldots,n) \qquad (5.14)$$

with "x" the elementwise matrix product.

To obtain the value of the minimum substitute $\hat{A},\hat{B}$, and $\hat{D}$ into (5.10):

$$\tau(\hat{A},\hat{B},\hat{D}) = \sum_{k=1}^{n} \text{tr } C_k'C_k - 2\,\text{tr } \hat{B}\hat{U}_B + \text{tr } \hat{B}\hat{V}_B\hat{B}'$$

using (5.13) this gives

$$\tau(\hat{A},\hat{B},\hat{D}) = \sum_{k=1}^{n} \text{tr } C_k'C_k - 2\,\text{tr } \hat{B}\hat{U}_B + \text{tr } \hat{B}\hat{V}_B\hat{V}_B^{-1}\hat{U}_B =$$

$$= \sum_{k=1}^{n} \text{tr } C_k'C_k - \text{tr } \hat{B}\hat{U}_B,$$

which with the analoguous result for (5.11) and (5.12) gives the desired result. ▼▼

Standardization of transformation matrices

Given a solution (A,B,D) has been found, any

$A^* = A\Delta$, $B^* = B\tilde{\Delta}$ and $D^* = (D_1^*, D_2^*, \dots, D^*_n)$ with

$D_k^* = \Delta^{-1} D_k \tilde{\Delta}^{-1}$ and $\Delta, \tilde{\Delta}$ full-rank diagonal matrices also

constitute a solution, as

$$A^* D_k^* B^{*\prime} = (A\Delta)\ (\Delta^{-1} D_k \tilde{\Delta}^{-1})\ \tilde{\Delta}B' = AD_kB'$$

Without restricting the generality we may, therefore, fix the arbitrary scaling constants δ, and $\tilde{\delta}$ in such a way that the columns of A and B have unit length (see also Carroll & Chang, 1970, p.288-289). Some such choice is necessary to identify the stationary equations, and this particular choice has the advantage that the orthonormal transformation procedure from the previous section is a special case of the non-singular one.

As mentioned before, the solution given above is the same as that given by Carroll & Chang (1970), only the subject weights d_p^k are here treated per plane. In other words, we present CANDECOMP here as a procedure for a component model with two reduced modes with a diagonal extended core matrix, rather than a procedure for a component model with three reduced modes with body diagonal core matrix (see the discussion of these models in Chapter 3).

Algorithm. Theorem 5.2 can be used to construct an alternating least squares algorithm to find the non-singular transformation A and B. One main iteration step of the NS-algorithm will be defined as:

A-substep

1. $UA_a = \sum_{k=1}^{n} D_k^{(a)} \; B_a' C_k'$
2. $VA_a = \sum_{k=1}^{n} D_k^{(a)} \; B_a' B_a D_k^{(a)}$
3. $\tilde{A} = UA_a' \; VA_a^{-1}$
4. $\Delta = \begin{pmatrix} \delta_1 & & \\ & \ddots & \\ & & \delta_s \end{pmatrix}$ with $\delta_p = \sqrt{\sum_{q=1}^{s}} \; \tilde{a}_{pq}$ $(p=1,\ldots,s)$
5. $A_{a+1} = \tilde{A}\Delta^{-1}$

B-substep

1. $UB_a = \sum_{k=1}^{n} D_k^{(a)} \; A_{a+1}' C_k$
2. $VB_a = \sum_{k=1}^{n} D_k^{(a)} \; A_{a+1}' A_{a+1} D_k^{(a)}$
3. $\tilde{B} = UB_a' \; VB_a^{-1}$
4. $\tilde{\Delta} = \begin{pmatrix} \tilde{\delta}_1 & & \\ & \ddots & \\ & & \tilde{\delta}_s \end{pmatrix}$ with $\tilde{\delta}_p = \sqrt{\sum_{q=1}^{s}} \; \tilde{b}_{pq}$ $(p=1,\ldots,s)$
5. $B_{a+1} = \tilde{B}\tilde{\Delta}^{-1}$

D-substep

2. $VD_a = (A_a' A_a) \times (B_a' B_a)$

   $UD_a = B_a' C_k' A_a$ $(k=1,\ldots,n)$

3. $d_{pk}^{(a+1)} = \sum_{q=1}^{s} (UD_a)_{qq} \; (VD_a^{-1})_{pq}$ $(p=1,\ldots,s;\; k=1,\ldots,n)$

5.4 COMPARISON OF TRANSFORMATION PROCEDURES

Once the optimal transformation matrices have been found, they can be applied to the component matrices with the inverse transformations applied to the core matrix. Thus, after having found the optimal orthonormal transformations $\hat{K}$ and $\hat{L}$, $\hat{Z}_k = GC_kH'$ may be decomposed as $\hat{Z}_k = G^*C_k^*H^{*'}$ with $G^* = G\hat{K}'$, $H^*=H\hat{L}'$ and $C_k^*=\hat{K}C_k\hat{L}'$. Similarly, after having found the optimal non-singular transformations $\hat{A}$ and $\hat{B}$, $\hat{Z}_k$ may be decomposed as $\hat{Z}_k = G^*C_k^*H^{*'}$ with $G^* = G\hat{A}$, $H^*=H\hat{B}$ and $C^*_k = \hat{A}^{-1}C_k(\hat{B}')^{-1}$. By using the transformations this way no additional loss is incurred and the transformed solution is just as adequate from the fitting point of view, as the original principal component solution.

In comparing the two transformation procedures it should be realized that the two loss functions have a different character, and that this leads to a number of differences in behaviour. The orthonormal loss function is

$$\sigma(K,L,D) = [[D_k - KC_kL']]^2 \quad , \text{K, L orthonormal}$$

which leads to $\hat{D}_k = \text{diag}\ [\hat{K}C_k\hat{L}']$, thus $\sigma(\hat{K},\hat{L},\hat{D})$ is the sum of squares of the off-diagonal elements of the transformed $\hat{C}_k^* = \hat{K}C_k\hat{L}'$, and because of that, ON is a true diagonalization procedure. The analoguous loss function in the non-singular case would be something like

$$\tilde{\tau}(A,B,D) = [[D_k - AC_kB']]^2 \quad , \text{A, B non-singular with unit length columns.}$$

The problem with this loss function is that the restrictions do not identify the minimization problem. A stricter requirement would be that the determinants of A and B are equal to one, but this could lead to rather complicated algorithms, which still have to be investigated.

The CANDECOMP loss function and solution described in section 5.3 was chosen for our preliminary investigations into non-singular transformations of the core matrix

$$\tau(A,B,D) = [[C_k - AD_kB']]^2 \quad \text{A, B non-singular with unit length columns.}$$

Properly speaking this is not a diagonalization procedure, but a decomposition of the core planes into the transformation matrices and diagonal matrices. $\hat{D}_k$ is not the diagonal of $C_k^* = A^{-1}C_k(B')^{-1}=$

when there is no exact solution, but the difference will become smaller when the loss becomes smaller. The parallel of the non-singular loss function for the orthonormal case would be the loss function for an *orthonormal INDSCAL* model, and the difference with the orthonormal transformation procedure in section 5.2 is that the definition of U and V is slightly different

$$U = \sum_{k=1}^{n} C_k L' D_k ,$$

and

$$V = \sum_{k=1}^{n} C_k' K' D_k .$$

For the non-singular case the difference in loss function implies that there is a difference between (1) the results (in terms of sums of squares) from the transformation procedure to find A and B, and (2) the results from applying A and B to the core matrix:

$$(1) \quad \text{NS} : \quad Z_k \cong G C_k H' \cong G\{A D_k B'\} H' = (GA) D_k (HB)'$$

$$(2) \quad \text{NS} : \quad Z_k \cong G C_k H' = (GA)\{A^{-1} C_k (B')^{-1}\}(HB)'$$

For comparison in the orthonormal case these decompositions are:

$$(1) \quad \text{ON} : \quad Z_k \cong G C_k H' \cong G\ \{K' D_k L\}\ H' = (GK') D_k (HL')'$$

$$(2) \quad \text{ON} : \quad Z_k \cong G C_k H' = (GK')\{K C_k L'\}(HL').$$

We will illustrate the differences between the two kinds of results for the non-singular procedure in section 5.5.

For interpretational purposes, there is a distinct disadvantage connected with the non-singular transformation when there is no exact solution to the diagonalization procedure NS. If there is no exact solution, $A^{-1} C_k (B')^{-1}$ is not diagonal, and the off-diagonal elements are no longer the sole expression of the relationships between components. Part of these relationships has been transferred to the non-orthogonality of the components themselves. No such complications occur with orthonormal transformations. In other words, the strength of the relationships between the components is now divided over two quantities, and this poses as yet unsolved interpretational complications. Especially in those cases where the non-singular transformations become nearly singular as, for instance, in the *Perceived reality study* (see section 5.5).

If the solution of the non-singular transformation is not exact, and one wants to avoid the complications of this splitting up of dependencies between components, one could settle for the extra loss from the transformation procedure, and use the values of D_k as the saliences or subject weights in combination with the correlated components $G^* = GA$, and $H^* = HB$, which are the same for (1) and (2) anyway. The scalar products $G^{*\prime}G^*$ and $H^{*\prime}H^*$ which are equal to $A'A$ and $B'B$ respectively then indicate the covariations of the components, or correlations if the component matrix is centred; the singular values of scalar product matrices will indicate the degree of non-singularity.

In section 5.5 we present some results of transforming the core matrix both by orthonormal and non-singular transformations.

5.5 ILLUSTRATIONS OF TRANSFORMATIONS

The main function of this section is to show numerical results illustrating the theory. Proper interpretation and assessment, especially of the non-singular transformation procedure, requires a more extensive investigation. For the orthonormal transformation procedure the situation is simpler, as it is a true diagonalization procedure, and the orthonormality of the transformation leaves the main characteristics of the TUCKALS2 solution unimpaired, and, therefore, poses no additional interpretational problems.

Four ability-factor study. (Meyers, Dingman, Orpet, Sitkei, & Watts, 1964; see Chapter 12) From Tables 5.1 and 5.2, which show the various core matrices, it follows that the orthonormal transformation of the core matrix does not improve the diagonality very much. Similar small improvements of diagonality can be observed in many other data sets, especially in those with symmetric frontal planes in the data block, like correlation and (dis)similarity matrices.

The non-singular transformation procedure succeeds rather well with a mere increase in the standardized loss of .0064 compared to the TUCKALS2 loss. This means that it is possible to decompose the

Table 5.1 *Four ability-factor study: results of transformation procedures (4x4 solution)*

Standardized sums of squares TUCKALS2 algorithm

Total sum of squares	-SS(Total)	1.0000
Fitted sum of squares (=sum of squares of core matrix)	-SS(Core)	.9394
Sum of squares of diagonal elements of core matrix	-SS(Dia)	.9247
Sum of squares of off-diagonal elements	-SS(Off)	.0147

Standardized sums of squares of transformation procedure

		ON	NS
Fitted sum of squares of the transformation procedure	-SS(Proc-Fit)	.9250	.9333
Sum of squares of the diagonal matrices D_k	-SS(Dia)	.9250	.6652

Standardized sums of squares after applying transformations

		ON	NS
Sum of squares of core matrix	-SS(Core)	.9394	.6922
Sum of squares of diagonal elements of core matrix	-SS(Dia)	.9250	.6749
Sum of squares of off-diagonal elements of core matrix	-SS(Off)	.0143	.0172

Transformation matrices

	Orthonormal case (K=L)				non-singular case (A=B)			
	$\boxed{1.000}$	.002	-.002	-.002	$\boxed{.972}$	-.189	.464	$\boxed{.841}$
	-.001	$\boxed{.996}$	.091	-.003	-.012	$\boxed{.926}$	-.349	.238
	.002	-.090	$\boxed{.983}$	-.161	-.207	.060	$\boxed{.646}$	.330
	.002	-.012	.161	$\boxed{.987}$	-.109	-.321	-.496	.357
Singular values of non-singular transformation					1.423	1.009	.814	.528

Average core matrices

three-mode analysis				orthonormal				non-singular			
$\boxed{10.79}$	-.04	.10	.03	$\boxed{10.79}$	-.04	.11	.06	$\boxed{7.99}$	-.01	.01	-.11
-.04	$\boxed{2.45}$	.03	.01	-.04	$\boxed{2.45}$	-.04	.01	-.01	$\boxed{2.32}$	.03	-.01
.10	.03	$\boxed{1.65}$	-.03	.11	-.04	$\boxed{1.66}$	.02	.01	.03	$\boxed{2.04}$	.00
.03	.02	-.03	$\boxed{1.37}$	.06	.01	.01	$\boxed{1.37}$	-.11	-.01	.00	$\boxed{4.08}$

frontal planes C_k of the core matrix into $\hat{A}\hat{D}_k\hat{B}'$ (or $\hat{A}\hat{D}_k\hat{A}'$ for the Meyers et al. data, as the input frontal planes are symmetric) without any real loss compared to the original TUCKALS2 solution. Thus the diagonal elements (see column 4, Table 5.2) can be interpreted as the weights or saliences which the groups attach to the axes of the common transformed space.

If one, in general, does not want to accept the additional loss, the core plane C_k should be transformed into

$$C_k^* = \hat{A}^{-1}C_k(\hat{B}')^{-1}$$

Table 5.2 *Four ability-factor study: core matrices* (x 10)

	TUCKALS2				Orthonormal				Non-singular				Diagonals from NS transformation procedure			
R2	79				79				65				56			
	-3	33			-3	33			- 4	28				30		
	2	-2	26		2	-3	26		-12	-4	29				24	
	7	3	1	13	7	2	1	13	-10	3	2	48				38
R4	135				135				98				103			
	3	20			3	20			-1	15				17		
	-4	-0	15		-4	-1	15		-3	-3	15				15	
	2	3	2	9	2	3	2	10	6	2	1	42				48
R6	116				116				65				75			
	4	32			4	33			3	35				33		
	7	4	8		6	2	7		9	2	11				13	
	6	-5	2	17	7	-4	1	18	7	-2	-4	52				57
N2	136				136				122				112			
	1	10			2	10			-2	10				9		
	-11	1	19		-10	1	19		-3	4	20				19	
	-8	1	-1	11	-10	1	-0	11	-2	-1	3	27				26
N4	84				84				56				55			
	-5	26			-4	26			-3	22				23		
	9	1	16		10	-0	16		8	0	25				29	
	-3	3	-2	15	-2	3	-2	15	-6	2	-1	41				35
N6	98				98				72				71			
	0	26			1	25			6	29				26		
	3	-1	16		3	-2	16		2	3	22				22	
	-2	-4	-2	16	-1	-5	-1	16	-1	5	-0	35				35

as explained in sections 5.3 and 5.4. When the fit is not perfect the C_k^* will not be completely diagonal, and the diagonal elements of C_k^* will not be exactly the same as in D_k (compare columnns 3 and 4 in Table 5.2), but the transformed component space $G^*=GA$ and $H^*=HB$ are the same as was set forth in section 5.4. The transformed components will in general no longer be orthogonal, and from Table 5.3 it can be seen that substantial correlations (and scalar products) may arise.

Table 5.3 shows the transformed components for Meyers et al. data. Whether one prefers the transformed components or the original ones, seems largely a matter of taste. The insight in the spatial arrangement of the tests (see Fig. 12.1) is not greatly enhanced by the non-singular transformation. On the other hand, the comparisons between the various weights the groups attach to the components is somewhat simpler due to the (near)diagonality. One

Table 5.3 *Four ability-factor study: transformed component space* (x 100)

Tests		1	2	3	4
A	1	32	35	7	14
Hand-Eye	2	30	38	-10	23
Psychomotor	3	32	41	-14	15
B	4	22	3	49	40
Perceptual	5	22	-1	48	43
Speed	6	22	-0	44	40
C	7	35	-39	25	3
Linguistic	8	32	-37	33	-3
Ability	9	34	-30	18	-3
D	10	22	-29	- 9	41
Figural	11	29	-30	-20	32
Reasoning	12	28	- 8	-15	34

component correlations

100			
- 3	100		
-27	-16	100	
-91	18	0	100

simpler due to the (near)diagonality. One of the reasons for the relatively small differences is that the core matrix was already reasonably diagonal to start with.

Perceived reality study. Non-singular transformations of the core matrix of this study (Van der Voort, 1982), discussed in detail in section 7.5, show an entirely different picture (Table 5.4).

Although the fit of the non-singular transformation procedure was quite good (the additional loss was only .0050), the results are far from attractive. The smallest singular values of the transformation matrices A and B are getting rather small, indicating that A and B are approaching singularity. Also noteworthy is the very large sum of squares of the core matrix. Note that the sum of squares of the core elements no longer adds up to the TUCKALS SS(Fit), because of correlations between components. The higher values in the core matrix after the non-singular transformation are the immediate consequence of these high correlations. It is, by the way, possible to scale the sum of squares of the transformed

Table 5.4 *Perceived reality study: results of transformation procedures*

(3x3 solution)

Standardized sums of squares

TUCKALS2		Transformation procedures	ON	NS	Application of transformations	ON	NS
SS(Tot)	1.0000	SS(Proc.Fit)	.8538	.8966	SS(Core)	.9016	5.1259
SS(Core)	.9016	SS(Dia)	.8538	4.6836	SS(Dia)	.8538	4.8062
SS(Dia)	.8468				SS(Off)	.0478	.3197
SS(Off)	.0548	iterations	22	> 200			

Transformation matrices

Orthonormal						Non-singular					
K			L			A			B		
$\boxed{.990}$	-.142	-.012	$\boxed{.996}$	.087	.006	$\boxed{.995}$	.490	$\boxed{-.942}$	.445	-.087	.030
.142	$\boxed{.990}$	.011	-.087	$\boxed{.992}$	.089	-.093	$\boxed{.727}$	.108	$\boxed{.828}$	$\boxed{.995}$	$\boxed{.884}$
.010	-.012	$\boxed{1.000}$	.002	-.090	$\boxed{.996}$	-.031	.482	.318	-.341	.052	.467
			singular values			1.46	.92	.16	1.63	.48	.31

Average core matrices

TUCKALS2			Orthonormal			Non-singular		
$\boxed{-7.27}$	.17	.05	$\boxed{-7.25}$	.23	.01	$\boxed{-15.14}$	.22	.28
.32	$\boxed{4.09}$	.31	- .36	$\boxed{4.13}$	-.04	.02	$\boxed{5.56}$	-.05
.09	-.27	$\boxed{1.77}$	- .00	-.16	$\boxed{1.78}$	.01	.21	$\boxed{12.17}$

core matrix down to its original size, but only by multiplying the components with the reciprocal scaling constants. Inspection of the transformed core matrix (not shown) indicates that large off-diagonal elements exist, thus pointing to non-diagonality. On the other hand, the good fit of the procedure shows that the D_k can be used as saliences for the transformed components. An adequate way to deal with this seeming contradiction still has to be developed.

5.6 CONCLUDING REMARKS

This chapter has been concerned with the problem of diagonality of the extended core matrix. Especially the non-singular transformation procedure is still problematic both technically and interpretationally. From a technical point of view it is not clear if the CANDECOMP procedure is the most adequate procedure to use for the purpose, and how diagonality should be measured in the presence of near-singularity of the transformation matrices. On the interpretational side the problem exists how to deal with large off-diagonal elements in a situation of good fit, and how to use the diagonality to its utmost advantage. The interpretation of highly related components is also somewhat difficult to deal with. The problem of interpreting 'oblique' components is, of course, partly conceptual and has been discussed extensively in the context of standard principal component analysis and factor analysis, a discussion we do not go into here.

Further insight into the behaviour of the solutions and further evaluation of the results may be obtained by a direct comparison with the results of, for instance, an INDSCAL analysis on the data of the *Four ability-factor study*, and a CANDECOMP analysis on the data of the *Perceived reality study*. This, however, merits another study.

II

THEORY
FOR
APPLICATIONS

SUMMARY

As Part I, *Part II* deals with theoretical issues, but now the focus is on those theoretical problems which arise out of applying three-mode principal component analysis to real data sets. Three main issues are tackled: preprocessing of input, postprocessing of output, and the analysis of the not-fitted part of the data.

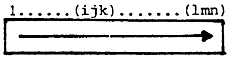
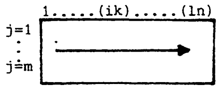
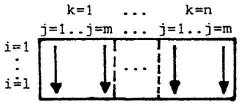
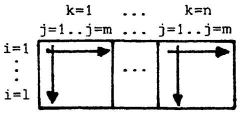
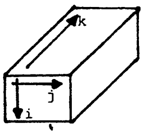
The first part of *Chapter 6* reviews proposals which have been put forward to scale input data, such that they are fit for a (three-mode) principal component analysis. Procedures for handling means and variances are discussed. To this end a distinction is made between (un)interpretable and (in)comparable means and variances. A large variety of models exist for dealing with interpretable means, which generally consist of additive terms for the means (in an analysis-of-variance fashion), and multiplicative or product terms for the components. Some such models are discussed and evaluated. The problem of iterative standardization in three-mode models is discussed briefly.

The second part of Chapter 6 deals with the interpretation of output, and ways to improve the interpretability of the results. Within this context the scaling of components and core matrices as well as their interpretation, joint plots, and component scores are treated in some detail.

In *Chapter* 7 the focus is on that part of the data which is not accommodated by the three-mode model. After a general discussion of residuals from principal component analysis, detailed recommendations and procedures are provided (and applied) for three-mode residuals, such as analysis of variance of the squared residuals, sums-of-squares plots, and the use of normal probability plots for the residuals.

SCALING AND INTERPRETATION

6

	type of centring	possibilities	data arrangement for centring
1.	none		
2.	over all data points z_{ijk}	*overall centring*	
3.	per element of one mode over the other two modes	i-centring *j-centring* k-centring	
4.	for each combination of elements of two modes over the elements of the third mode	ij-centring *(abative)* *jk-centring* *(ipsative)* ki-centring *(normative)*	
5.	per element of one mode over the elements of the other two mode separately: double-centring	ij,jk-centring *jk,ki-centring* *(performative* ki,ij-centring	
6.	triple-centring		

6.1 INTRODUCTION

In this chapter we will discuss the input and output of three-mode principal component analysis. First we will treat general issues in connection with means and variances of raw data, their influence on the analysis, and their treatment so as to obtain the 'best' analysis for a particular data set. Then we turn to what comes out of a three-mode analysis, how to interpret it, and how to transform it to enhance interpretation. In this sense the present chapter is the theoretical counterpart of the discussion in the second part of Chapter 2, in which we described an example in detail. A similarly detailed analysis of an example can be found in Chapter 8.

The first part of this chapter deals with scaling of *input*, and will lead us to consider mixed additive and multiplicative models for raw data, i.e. models which have properties of both analysis of variance and principal component analysis. It is also necessary to consider the purposes of *input scaling*. By *scaling* we mean any operation which transforms raw data into new data values by subtracting and/or dividing the former by certain, often data dependent, quantities, such as means, scale midpoints, standard deviations, ranges, etc.

In the second part scaling of *output* is considered. The Tucker2 and Tucker3 models are liable to what is sometimes called 'the fundamental indeterminacy' (e.g. Kruskal, 1981, p.5), i.e. the component matrices may be transformed non-singularly without changing the fit of the model to the data, provided the appropriate inverse transformations are applied to other parts of the model. Similarly, component matrices may be multiplied or divided by constants without affecting the fit of the model. Some such scalings of the output, however, enhance our understanding of the relationships underlying the data more than others, as we shall presently see.

6.2 INPUT SCALING: GENERAL CONSIDERATIONS

Types of scaling. We will primarily discuss two basic kinds of scaling: *centring,* i.e. "subtracting a constant term from every element so that resulting data values have a mean 0" (Kruskal, 1981, p.15), and *standardization,* i.e. "dividing every element by a constant term, so as to achieve this result: the 'scale' of the resulting data values has some fixed value (often chosen to be 1). 'Scale' generally refers to some measures of variability, most often the standard deviation" (Kruskal, 1981, p.17). The process of centring and standardization, such that the resulting data values have mean zero and standard deviation one will be called *normalization.*

Although we primarily look at these three operations, it does not always seem adviseable to revert to them. In the examples in Chapters 9 and 11 we have subtracted the scale midpoints from the data for reasons to be explained later. In Chapter 14 we adjusted the range of all variables (tests) to become identical.

Selecting a type of scaling. The reasons why one should use a particular type of scaling depend on the position one takes with respect to data and their analysis. The first point of view is that measurement characteristics, research questions, and research design determine what ought to be done to the data before entering a three-mode principal component analysis. A second point of view is that the model determines also which kind of scaling or preprocessing is appropriate, i.e. the model rules out certain scalings, as they are considered to be inconsistent with its definition.

First, it is necessary to take a closer look at the question why one should need to consider scaling at all. The answer to this is straightforward: to improve the understanding of the relationships between elements of the three modes. It is believed that the three-mode component model does not apply to the raw data, but to some appropriately transformed or scaled form, and that certain means and/or standard deviations obscure what is searched for, or what is basic in the data. Thus it is expected that the model will give an incorrect or imprecise description of the relationships in the data, when applied to the raw data.

Why one might get an improper description when certain means and/of standard deviations are not removed from the raw data follows from the definition of component analysis. The components derived by the technique represent directions in the space spanned by, say, the variables, along which successively and orthogonally the largest variations can be found. If the centroid, defined by the means of the variables, is located at a considerable distance from the origin of the variable space, then an important candidate for the direction of the first component will be the one from the origin through the centroid. If, however, the main purpose of an analysis is to investigate the covariations of the variables from the centroid, the means of the variables should be removed before the component analysis, and should be modelled separately. Similarly, when the structure of the variable domain is of interest, but it is undesirable that variables with larger variations influence the results unduly, something should be done towards equalizing variations. It is equally possible, however, that objects or persons with larger variations should dominate the outcome of an analysis. For instance, it is not necessarily sensible to equalize variations of persons who have no outspoken opinion and always tick the midpoints of scales and those of persons who use scales effectively and have mainly systematic variation.

Kruskal (1981, p.18) cites another purpose for standardization in connection with his discussion of PARAFAC1. Paraphrasing his argument we write the three-mode Tucker3 model as

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

where ε_{ijk} is a random variable with mean zero. As least squares fitting is used to solve the estimation of the model, it is implicitly assumed that the standard deviations of the error terms are all equal. If one knew these standard deviations, one could scale the z_{ijk} to make the error terms as nearly equal as possible. As in practice one does not know the standard deviations of the error terms, one has to fall back on the idea frequently used in principal component analysis, i.e. seeking to make the total standard deviations of the elements of one of the modes equal [or a similar type of standardization] instead of seeking to make the error standard deviations equal. "This approach has a long tradition in the bilinear methods, and is presumably as reasonable for trilinear models [such as PARAFAC1] as bilinear models [such as principal component analysis], though a satisfying rationale for it is not known" (Kruskal, 1981, p.18).

Returning to the two points of view to input scaling, both Kruskal (1981) and Harshman (cited in Kruskal) argue that certain scalings are inappropriate for the three-mode model as the components after centring and/or standardization bear no simple relation to the components before transformation. Put differently: a scaling should not "destroy the agreement with the model" (Kruskal, 1981, p.18). As it is our contention that the principal component model generally only applies *after* transformation of the data values it is not necessary to compare components before and after transformation. We cannot go into the question here in more detail, as little discussion of this issue has appeared in print with respect to three-mode models. The basic paper seems to be an informal and incomplete paper by Harshman (cited in Kruskal, 1981). A final draft is in preparation (Harshman, 1982, pers. comm.), and due to appear in Law, Snyder, Hattie, & McDonald (forthcoming).

Types of three-mode data. In selecting an appropriate scaling it is important to distinguish between three general kinds of three-mode data which we will designate as 'principal component analysis data' or *pca-data*, 'multidimensional scaling data' or *mds-data*, and 'analysis of variance data' or *anova-data*.

Pca-data have the format: subjects (i-mode) × variables (j-mode) × conditions (k-mode). The terms are generic ones, e.g. conditions may refer to points in time, occasions, experimental conditions, replications, etc. The subjects may be considered a (random) sample from a particular population, or a fixed group of persons about which information on individual differences is sought. The examples in Chapters 8 *(Attachment study)*, 13 *(Hospital study)*, and 14 *(Learning-to-read study)* have this data format.

Mds-data have the format: variables, stimuli, or scales (i-mode) × variables or stimuli (j-mode) × subjects (k-mode). Characteristic of this kind of data is that the subjects are not considered mere replications, but nearly always their individual differences are of interest, and they are seldom treated statistically as if they were a (random) sample from a particular population. The examples in Chapters 2 *(Party similarity study)*, 9 *(Triple personality study)*, 10 *(ITP study)*, 11 *(Cola study)*, and 12 *(Four ability-factor study)* have this data format.

The third, not too common type of data (*anova-data*), generally have the pca-format with the additional characteristic that the variables (j-mode) form a highly consistent scale (high Cronbach's α) and may be considered to measure the same variable. In such a case the data may be described by a three-factor (ℓ×m×n) analysis of variance design without replications. The *Perceived reality* data in Chapter 7 have this data format.

As in any classification scheme the allocation to one of the formats is not always clear-cut. In fact, the *Triple personality data* of Chapter 9 could be considered both pca-, and mds-data, but treating them as mds-data seems to be more in line with the research questions asked. Four-mode data will in many cases be mixtures of the pca- and mds-data, see e.g. the data collected by Jones & Young (1972), when the two years in which mds-data were collected, are considered as the fourth mode. The distinction between pca-data and mds-data is especially useful in connection with the decisions which of the modes should not be reduced in a Tucker2 analysis. For the former this will be the condition mode, for the latter the subject mode. It is, by the way, interesting to note that the substantive distinction between the mathematically equivalent models PARAFAC1 and CANDECOMP (see section 3.2 and section 3.3) is that the former was proposed with pca-data in mind, and the latter with mds-data.

In section 6.5 we discuss recommendations for centring and will return to these three data formats.

6.3 INPUT SCALING: ARBITRARY AND INCOMPARABLE MEANS AND VARIANCES

Arbitrary means and variances. Many social science variables have interval properties, and thus no natural zero point. Often the absolute size of the variances of these values is arbitrary in the sense that it is dependent on the number of possible values chosen rather than the 'true' range of the variable. It is undesirable to have variation in the data due to arbitrary means (e.g. the midpoint of five-point rating scales) influence the components of a component analysis, so that they should be removed first - the more so if they are different for different variables. In certain cases with homogeneous variables (for instance, sets of similar rating scales) the differences in the arbitrary means are of interest, and should be retained in the analysis. In that case the midpoints of the variables generally define some neutral point which can be used for centring (see the examples in Chapters 9 and 11).

The situation with variances is similar. If variances are arbitrary, and variables have different ranges, then in order to avoid artefacts these variances should certainly be equalized before a three-mode analysis is performed. In homogeneous sets with arbitrary variances, in which the differences between the variances are not of interest, they should be equalized as well. When the differences are of interest, the variances probably should remain untouched, as standardization per variable will remove some or all of these differences. One could consider scaling the overall variance to unity over the entire data set, but this has no influence on the outcome of the analysis because all the data values are divided by the same constant. (see also Kruskal, 1981, p.17).

Incomparable means and variances. Consider the situation in which the scores of a number of subjects are available on a diverse collection of variables, each of which has its own measurement characteristics. The *Hospital study* in Chapter 13 may serve as an example: 188 hospitals were measured on variables like *number of beds, presence or absence of a financial director, ratio of qualified nurses to the total number of nurses,* etc. In such data the means of the variables are incomparable, as are the variances. Therefore, it does not make sense to consider components which are influenced by these means and variances. In other words, these means and variances should be modelled separately, and not via a principal component analysis.

The hospital data are, in fact, more complex than sketched above, because the variables were measured in each of eleven consecutive years. The question thus arises whether one wants to remove the incomparable means per year, or over all years together. The argument for the scaling procedure in the previous paragraph was based on the idea that within one year the means and standard deviations across variables were incomparable. However, differences in means and standard deviations over the years are comparable for each variable, and one may decide to model the differences across years by the principal component analysis, or model them separately outside the model, depending on the research questions one has in mind.

In this way one may have both incomparable and comparable means in one data set, and the 'best' way to treat them depends on one's view of the subject matter. It may, of course, happen that one has to perform more than one kind of scaling due to lack of insight in the data set itself.

In other situations all means and/or variances are both interpretable and comparable, e.g. all variables are bipolar scales as in semantic differential research (see Chapter 9). It is then a question whether means and/or variances should be modelled separately or not.

6.4 INPUT CENTRING: INTERPRETABLE MEANS

In this section we will investigate some of the substantive considerations that go into selecting an appropriate centring for a particular data set when the means are *interpretable*. It is not possible to do so in a general way because different research questions are asked of different data, and because different measuring instruments and research designs are in use. It is possible to make specific recommendations in specific research areas, as has been demonstrated by Noy-Meir (1973), and Noy-Meir, Walker, & Williams (1975) for ecological ordination data.

Notwithstanding the above we will try to tackle centring of input data as generally as possible by discussing various ways in which means can be treated and/or modelled. Our emphasis will be primarily on centring as this kind of scaling is better understood, and more extensively studied. We will discuss standardization in some more detail in section 6.6.

Two-mode data. To facilitate the discussion let us assume that we are dealing with scores of individuals on a series of tests scored on the same scale. The means of these tests are comparable, as are those of the individuals. Assuming that it makes sense to talk about the average performance of an individual over all tests, the question arises as to how the average performance should be modelled. Similarly, given that we have determined the averages of all tests, the question arises how they should be included in an analysis. One way to do so is to perform a standard principal component analysis, or singular value decomposition (see section 2.2) on the original measures.

An alternative way to treat these means (and the means of the individuals over tests) would be to model them according to a model sometimes called the FANOVA (FActor ANalysis Of VAriance) model (Gollob, 1968a,b,c). This model treats the grand mean, row and column effects separately, i.e. removes them from the original data, and specifies a singular value decomposition for the residuals. The derived components are 'interaction-components' in that they describe the interactions of the deviations from the means of the individuals and the tests respectively:

$$z_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \text{ with } \varepsilon_{ij} = \sum_{p=1}^{s} g_{ip} h_{jp} c_{pp},$$

where μ, α_i, β_j, ε_{ij} are the usual grand mean, row effect, column effect, and residual from analysis-of-variance models, with the standard zero-sum assumptions for the effects (see also Kruskal, 1981, p.6,7). One hopes, of course, that very few components are necessary to describe the interactions. The FANOVA model is thus a combination of an additive model (grand mean, row effect, column effect), and a multiplicative model (componental decomposition of the remainder term). The latter part is, however, still an orthogonal decomposition, and in that sense the successive product terms are also additive between them. One should also realize that the model as specified is one without replication, i.e. with only one observation per cell.

The main differences between the two ways of modelling, FANOVA and singular value decomposition, are the treatment of the means and the interpretational differences connected with the components. Tucker (1968), for instance, contends that "the mean responses to various stimuli over a population of individuals are better conceived as describers of the population than as simple, fundamental describers of the stimuli" (p. 345), and continues that, therefore, such means should be included in a principal component analysis, i.e. the analysis should be performed on the original measures. In this way the means are "equal to the measures that would be observed for a person at the centroid of the factor score distribution" (p. 350). The components then determine the original measures.

In contrast, the FANOVA model sets the means apart first, and only then looks at components in the residuals. It, therefore, gives a special a priori status to those means. It is a moot point whether this is just "a useful heuristic to use main effects as a point of reference from which to describe individual differences in patterns of subject responses" (Gollob, 1968c, p. 355), or whether in the FANOVA model "the mean measure is considered as a basic characteristic of the responses of the individuals" (Tucker, 1968, p. 350). In the end the subject matter will determine which of the two is the more correct interpretation of the mean responses, and the research questions will determine if it is more useful to model the means a priori (Gollob) or a posteriori (Tucker). When the means are expected to be the resultant of an agglomeration of influences which have to be disentangled, Tucker's view seems to be pertinent. However, when the means represent a 'primary psychological construct', or have intrinsic meaning in another way Gollob's view and model seem more appropriate.

Whereas Gollob and Tucker discuss the FANOVA model within the context of the problems of removing or maintaining means before performing a principal component analysis, the same model has been considered from a different angle by Mandel (1969, 1971), and in fact even earlier by Fisher & Mackenzie (1923) and Gilbert (1963). Mandel was looking for a way to model the interactions in a two-way analysis-of-variance design without replications, and attempted to fit multiplicative interactions of row and column factors, ending up with the same model as Gollob. He thereby extended the already existing discussion on tests for non-additivity which started with Tukey's (1949) 'single-degree-of-freedom test for non-additivity'. Further work on testing this kind of interactions was carried out by Corsten & Van Eijnsbergen (1972), Johnson & Graybill (1972), and Marasinghe & Johnson (1981). Snee (1972) discusses the model for growth studies (see also Chapter 14).

Notice that within this context there is no problem as to whether or not it is appropriate to remove means, as the primary focus is on modelling interaction terms after the sums of squares for the main effects have already been investigated. Another and more fundamental difference between the two presentations of the model is the kind of data involved. Whereas Gollob considers observations of subjects on certain variables, and therefore looks at the relationships between variables (analysis of interdependence, see e.g. Gifi, 1981, p. 2., and Kendall, 1957, p. 1-4), Mandel is dealing with one response variable and two predictor variables (analysis of dependence). Because of this fundamental difference not all considerations, tests, etc. from the analysis-of-variance-side are relevant to the Gollob-Tucker discussion, and vice versa.

A parallel model crops up in the analysis of contingency tables, where it goes under a wide variety of names, such as correspondence analysis (see Chapter 15), dual scaling, and optimal scaling. For general surveys, historical comments and extensive references see Hill (1974), Gifi (1981) and Nishisato (1981).

Gabriel (1971) used the FANOVA model for so-called biplot graphical analysis of multivariate data, and Gnanadesikan & Kettenring (1972, p.97,102) also implicitly suggest the use of the model when discussing ways to investigate residuals for outliers (see also Chapter 7).

Reviewing the various discussions of the model in the above papers as far as they are relevant to three-mode models, it seems that the crucial aspects are the kind of research questions being asked, and the research design used to collect the data. This should determine what to do with the row and column means, or main effects, be it that it is often far from easy in practical cases to decide upon the proper way of centring. Only after this matter is solved, one can turn to a multiplicative analysis of interactions by using singular value decomposition, or its three-mode analogues such as three-mode principal component analysis or simplified versions thereof (see Chapter 3). In the next subsection we will review some three-mode generalizations of the FANOVA model.

Three-mode data. Lohmöller (1979) discusses additive and multiplicative models for three-mode data, including some that fall outside the present discussion. He suggests the following generalization of the FANOVA model:

$$z_{ijk} = \mu + \beta_j + \gamma_k + \zeta_{jk} + \sigma_{jk} \times \tilde{z}_{ijk}$$

with the normal analysis-of-variance notation for the grand mean (μ), the variable effect (β_j), the condition effect (γ_k), and the combined variable/condition interaction effect (ζ_{jk}). The remaining jk-normalized $\tilde{z}_{ijk}$ are to be decomposed with the three-mode principal component model. This model (called the *standard reduction equation* by Lohmöller) specifies the data partly as an additive function of a priori sources of variation (μ,β,γ,ζ), with standardization constants σ, and partly as a posteriori sources of variation through the components. The discussion in the previous paragraph on the appropriateness of removing means before a multiplicative analysis directly applies to this proposal and the ones discussed below.

Once it is realized that analysing interpretable means separately implies nothing but an analysis-of-variance model with multiplicative interaction terms, there is a large number of models that may be proposed. One possibility is the three-way main effects analysis-of-variance model for the additive part, and the three-mode principal component model for the multiplicative part. The triple-centring model shown in Table 6.1 could also be used in this way, although it remains to be shown that it is a really useful procedure. It may very well be that after the various means and interactions have been removed, the residuals will not contain much additional systematic information that can be described by three-mode principal component analysis. The deviations from randomness in these remainders might be better investigated by some kind of residual analysis (see also Chapter 7).

Whereas in the above discussion three-mode models with multiplicative interactions have been approached from the component analysis side, Gower (1977) follows the analysis-of-variance tradition of Mandel. He describes three-way models which fit the overall mean and main effects additively, and two-way interactions multiplicatively:

$$z_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk}, \text{ with } \varepsilon_{ijk} = g_i h_j + \ddot{h}_j e_k + \ddot{e}_k \ddot{g}_i$$

with in general different components g_i, $\ddot{g}_i$ and $h_j, \ddot{h}_j$, and $e_k, \ddot{e}_k$, which are derived from separate singular value decompositions of the two-mode marginal matrices averaged over one subscript. It is assumed that all effects and multiplicative components sum to zero, and that there is no three-way interaction. The additive portion is fitted by the standard least squares estimators, and the multiplicative part is based on the residuals,

$$\varepsilon_{ijk} = z_{ijk} - z_{...} - (z_{i..} - z_{...}) - (z_{.j.} - z_{...}) - (z_{..k} - z_{...}).$$

Gower proposes to use Mandel's (1971) formulas on degrees of freedom in the two-way case to compute mean squares, and to test the significance. Note that in our discussion of Gower's and Mandel's models the assumption is made that there are no replications in the cells. When there are replications, their models would be four-mode models in our terminology with one random mode.

Gower continues to show that the inclusion of three-component products, for instance $g_i h_j e_k$, introduces further complications when one wants to include the zero-sum restrictions on the multiplicative components. Including such restrictions leads to a model like

$$z_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk} \text{ with}$$

$$\varepsilon_{ijk} = g_i h_j + h_j e_k + e_k g_i + \rho g_i h_j e_k$$

with zero-sum restrictions for all effects and multiplicative components; ρ is a constant to be estimated. Note that there is now only one type g_i, h_j, and e_k.An even more complicated three-factor model is considered when separate two-way interactions are included as well. Gower discusses estimation procedures for the above model and the difficulties involved.

The analysis-of-variance approach colours the way the model is conceived and the way restrictions are introduced. This is, for instance, evident in the insistence on zero-sum restrictions, and the inclusion of two-way interactions before introducing three-way interactions. The component analysis approach decomposes three-way interactions directly without necessarily fitting two-way interactions first. The two approaches coincide when p=1 and two-way interactions are ignored, i.e. $g_i h_j = g_i e_k = h_j e_k = 0$ for all i,j, and k. The remaining term has then exactly the form of the PARAFAC1/ CANDECOMP model (Harshman, 1970; Carroll & Chang, 1970; see also sections 3.2 and 3.3) with one or s components:

$$\varepsilon_{ijk} = g_i h_j e_k, \text{ or } \varepsilon_{ijk} = \sum_{p=1}^{s} g_{ip} h_{jp} e_{kp},$$

depending on the numbers of multiplicative terms one wants to include. Furthermore, when the two-way interactions again are not explicitly modelled, a three-mode principal component model is identical to Gower's model, when only one multiplicative term is included. The c_{pqr} with p=q=r=1 is then the estimator of ρ .

De Leeuw (1982, pers. comm.) suggested an extension of the Tucker3 model which bears some resemblance to the model with three-component products proposed by Gower, and at the same time solves the estimation problem via an alternating least squares algorithm.

Assume that all component matrices G, H, and E of a three-mode principal component model have a constant first column, i.e. $g_{i1} = 1/\sqrt{\ell}$ for all i, $h_{j1} = 1/\sqrt{m}$ for all j, $e_{k1} = 1/\sqrt{n}$ for all k. Then we may write this modified version of the basic Tucker3 model as

$$z_{ijk} = c_{111}/\sqrt{\ell mn} + (1/\sqrt{\ell m}) \sum_{r=2}^{u} e_{kr} c_{11r} + (1/\sqrt{\ell n}) \sum_{q=2}^{t} h_{jq} c_{1q1} +$$

$$(1/\sqrt{mn}) \sum_{p=2}^{s} g_{ip} c_{p11} + (1/\sqrt{\ell}) \sum_{q=2}^{t} \sum_{r=2}^{u} h_{jq} e_{kr} c_{1qr} +$$

$$(1/\sqrt{m}) \sum_{p=2}^{s} \sum_{r=2}^{u} g_{ip} e_{kr} c_{p1r} + (1/\sqrt{n}) \sum_{p=2}^{s} \sum_{q=2}^{t} g_{ip} h_{jq} c_{pq1} +$$

$$\sum_{p=2}^{s} \sum_{q=2}^{t} \sum_{r=2}^{u} g_{ip} h_{jq} e_{kr} c_{pqr} .$$

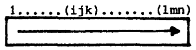
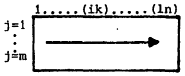
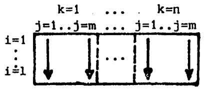
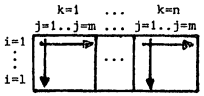
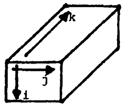
The PARAFAC1/CANDECOMP version of this modified Tucker3 model can be obtained by this setting all $c_{pqr} = 0$ except when p=q=r, and absorbing the constants and the c_{ppp} in the components:

$$z_{ijk} = c_{111} + e_{k1} + h_{j1} + g_{i1} + \sum_{p=2}^{s} h_{jp} e_{kp} + \sum_{p=2}^{s} g_{ip} e_{kp} +$$

$$\sum_{p=2}^{s} g_{ip} h_{jp} + \sum_{p=2}^{s} g_{ip} h_{jq} e_{kr},$$

and this is Gower's model if s=2. The estimation of this model can thus be solved by adapting the PARAFAC1/CANDECOMP algorithm. Similarly the estimation of the modified Tucker3 model can be solved by adapting the TUCKALS3 algorithm.

Table 6.1 *Types of centring three-mode data*

type of centring	possibilities[+]	for *selected* possibility				references[§]
		one-way means to be retained	eliminated means	formula	data arrangement[†] for centring	
1. none		all	none	$\tilde{z}_{ijk} = z_{ijk}$		Tucker(1966)-*a*;V.d.Geer(Note 4)-*1*; TUCKALS-*option 0*.
2. over all data points z_{ijk}	*overall centring*	all	$z_{...}$	$\tilde{z}_{ijk} = z_{ijk} - z_{...}$	 1......(ijk).......(lmn)	Bartussek(1973)-*a*;V.d.Geer-*2*; TUCKALS-*option 8*.
3. per element of one mode over the other two modes	i-centring *j-centring* k-centring	between subjects (i), between conditions (k)	$z_{.j.}$	$\tilde{z}_{ijk} = z_{ijk} - z_{.j.}$	 1.....(ik).....(ln) j=1 : j=m	*j- centring* Tucker-*b*;Bartussek-*b*; V.d.Geer-*3*. *k- centring:* TUCKALS-*option 4* .
4. for each combination of elements of two modes over the elements of the third mode	ij-centring *(abative)* *jk-centring* *(ipsative)* ki-centring *(normative)*	between subjects (i)	$z_{.jk}$	$\tilde{z}_{ijk} = z_{ijk} - z_{.jk}$	 k=1 ... k=n j=1..j=m ... j=1..j=m i=1 : i=1	*jk-centring* Tucker-*c*;Bartussek-*c*; V.d.Geer-*5*;Lohmöller(1979); TUCKALS-*option 2*. *ki-centring* Bartussek-*d*;TUCKALS-*option 3*.
5. per element of one mode over the elements of the other two mode separately: double-centring	ij,jk-centring *jk,ki-centring* *(performative)* ki,ij-centring	between conditions (k)	$z_{.jk}, z_{i.k}$	$\tilde{z}_{ijk} = z_{ijk} - z_{.jk} - z_{i.k} + z_{..k}$	 k=1 ... k=n j=1..j=m ... j=1..j=m i=1 : i=1	*general:* Gollob(1968b,c);Tucker(1968). *jk,ki-centring* TUCKALS-*option 1*.
6. triple-centring		none	$z_{.jk}, z_{i.k}$ $z_{ij.}$	$\tilde{z}_{ijk} = z_{ijk} - z_{.jk} - z_{i.k} - z_{ij.} + z_{i..} + z_{.j.} + z_{..k} - z_{...}$	 k j i	

[+] The type of centring in *italics* is the one worked out in the right hand part of the table. The *(italicized terms)* are Cattell's (1966a).

[†] The arrows indicate the direction of averaging.

[§] TUCKALS is the generic name for two computer programs TUCKALS2 and TUCKALS3 developed by Kroonenberg & De Leeuw (1980), Kroonenberg (1981a,c) Tucker (1966)-*a* means that this way of centring was labelled *a* by Tucker in his 1966 a paper

6.5 INPUT CENTRING: TYPES, CONSEQUENCES, RECOMMENDATIONS

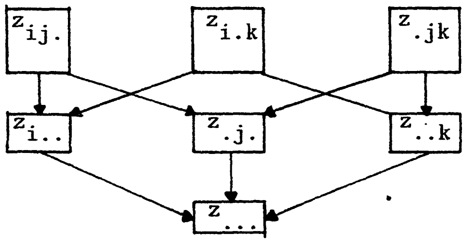
In the previous section we looked at substantive issues connected with input centring, and at some models which could be used for treating means separately from the component model. In this section we will look at centring from a more technical point of view by considering the kinds of centring which can be defined within three-mode analysis, and the effects these centrings have on the output. Finally, we will try to formulate some recommendations, as well as discuss those of others.

Types of centring. In Table 6.1 an overview is given of centring possibilities for three-mode data matrices. Cattell (1966a, p.115-119) has coined some terms for scaling of two-mode matrices. Whenever his terms seemed applicable for three-mode data, we have included them in italics in Table 6.1. Tucker (1966a, p.294), Bartussek (1973, p.180-181), Van de Geer (1975, p.12), Lohmöller (1979, p.156-158), Rowe (1979, p.78), Harshman (unpublished, quoted in Kruskal, 1981), and Kruskal (1981, p.15-19) discuss the scaling of input data for three-mode models, and most of the schemes for centring have been proposed by at least one of them.

Some consequences of centring. In Table 6.2 an overview of some consequences of centring with various schemes is given for both the Tucker3 and Tucker2 models. The general effect for means is that if centring takes place at a certain level, means at the lower connected levels will also be zero. Especially noteworthy is Case 5 (jk,ik-centring or double-centring) as the only non-zero means remaining are those in the $Z_{ij.}$ two-dimensional marginal plane (average frontal plane of the data matrix). Note, however, that its one-dimensional marginal averages are both zero again. In other words, the average frontal plane of the data matrix is itself double-centred. For the triple-centring indicated, the grand mean and all one-dimensional and two-dimensional marginal averages are zero.

Table 6.2 *Consequences of centring*

A.



two-dimensional marginal averages

one-dimensional marginal averages

overall (grand) mean

General rule: if means are removed at a level, they will also have been removed at a lower level.

B.

case	type means removed *type of centring*	consequences
2	$z_{...}$ *overall centring*	no components centred
3	$z_{.j.}$ *j-centring*	ik-combination-mode components centred (e.g. in Tucker's Method III) component scores on j-mode components centred
4	$z_{.jk}$ *jk-centring*	T3: i-mode components centred T2: (k or j-mode unreduced): i-mode components centred T2: (i-mode unreduced): c_{iqr} centred per qr-combination tu×tu latent covariation matrix becomes covariance matrix
5	$z_{.jk}$, $z_{i.k}$ *jk, ik-centring*	T3: i-mode and j-mode components centred T2: (k-mode unreduced): j-mode and i-mode components centred T2: (i-mode unreduced): j-mode components centred c_{iqr} centred per pr-combination c(tu×tu) latent covariance matrix T2: (j-mode unreduced): i-mode components centred c_{pjr} centred per pr-combination (su×su) latent covariance matrix

6	$z_{.jk}, z_{i.k}, z_{ij.}$	T3: i-mode, j-mode, k-mode components centred
	triple-centring	T2: components of reduced modes are centred. Core matrix is centred over unreduced mode. Latent covariance matrix for unreduced component combinations.

* cases as in Table 6.1

The last column of Table 6.2B indicates some effects of the various ways of centring on the output of a T3 or T2 analysis. Particularly, it shows which component matrices and core matrices become centred, and for which 'latent covariation matrices' the entries will become covariances (see section 13.3).

Uncentred modes will have large first components, due to the presence of the means. This should be taken into account when assessing the relative contributions of the components of such an uncentred component. The first components are often highly correlated with mean vectors, and also with the fitted sums of squares of the elements of the mode.

When per variable (j-mode) is centred for each condition of the k-mode (jk-centring) the i-mode components will be centred after analysis. The algebraic correctness thereof follows directly from:

$$\tilde{z}_{ijk} - \tilde{z}_{.jk} = \sum_{p,q,r} g_{ip}h_{jq}e_{kr}c_{pqr} - (1/\ell) \sum_{i} \sum_{p,q,r} g_{ip}h_{jq}e_{kr}c_{pqr} =$$

$$= \sum_{p,q,r} h_{jq}e_{kr}c_{pqr}\{g_{ip} - (1/\ell) \sum_{i} g_{ip}\} =$$

$$= \sum_{p,q,r} h_{jq}e_{kr}c_{pqr}\{g_{ip} - \bar{g}_{.p}\} ,$$

with $\tilde{z}_{ijk}$ the centred scores, such that $\tilde{z}_{.jk} = 0$ for all j and k. At the same time the algebraic derivation shows that it is more of a terminological confusion than anything else, as both by the raw data and in the model the centring is over the index i. In other words, one should keep track of the index *over* which is centred. Part of the confusion is due to our use of the word component, i.e. in our terminology components refer to loadings (see section 1.4), rather than component scores, as is more usual in standard principal component analysis. In standard principal component analysis the scores of subjects on the variables are deviation scores after centring, and so are the scores on the variable components (see also Figure 2.1).

Another problem which has been raised with respect to centring is the relationship between components derived from various centring procedures. An extensive literature exists dealing with two-mode matrices, but this discussion will not be repeated here. The presence of three modes and thus three component matrices makes the matter considerably more complex, and little work has been published concerning this problem (see Kruskal, 1981). Whether an investigation will be useful for practical applications is rather doubtful, since just as in the two-mode case it is substantive considerations which generally determine the kind of input scaling that will be used (see section 6.2). Some relevant references for the two-mode discussion are Harris (1953), Ross (1963, 1964), McDonald (1967, especially p. 8-10), Gollob (1968b,c), Tucker (1968), Corballis (1971), Nesselroade (1973), Noy-Meir (1973), and Noy-Meir, Walker, & Williams (1975).

The consequence of removing any mean is that the amount of variation explained by a particular analysis will be smaller, and sometimes dramatically smaller. (The sum of squares caused by non-zero means is often the largest one present in the data.) Overall centring (Case 2) can, for instance, be interpreted geometrically as moving the centroid of the data to the origin, and thus the sum of squares caused by the overall mean (which is often not meaningful, as in the case of many rating scales) is removed. Therefore, the loss in (artificial) variation explained need not be regretted in such a case.

A further problem in connection with centring has to do with the interaction between centring and standardization, which we will take up briefly in the next section.

Recommendations. The recommendations presented here should be seen as a first guide to what can be done with a particular data set. Especially in situations in which means can be modelled, much more content-specific information is needed.

When means cannot be interpreted or when they are incomparable within a mode, they should be eliminated, i.e. set equal to zero via subtraction. Furthermore, when means are interpretable and comparable within a mode, and when it is evident that the differences have been caused by influences not connected with the three-mode data itself, they had best be modelled and explained separately outside the three-mode model.

For pca-data (see section 6.2) one will generally use either j-centring or jk-centring. The choice between the two will mainly depend upon the evaluation of the causes of the differences in means across conditions, and the need to relate these means to other aspects of the data set.

For mds-data the most common procedure is double-centring (jk, ik-centring). Since the subjects in the third mode are assumed to be independent and we want to describe individual differences in the way the stimuli (variables) are treated by the subjects, the data should be centred per subject, or matrix-conditional (see, e.g. Young, 1981). In the literature on multidimensional scaling (see, e.g. Torgerson, 1958, p.258), the data values, which are assumed to have distance-like properties, are often first squared before double-centring, so that the double-centred values, $\tilde{z}_{ijk}$, can be interpreted as scalar products between i and j for each k. Alternatively one could consider the observed values as being already squared, as we have generally done in our examples (Chapter 10 and 11), and as Tucker (1972) did to demonstrate three-mode scaling. One of the effects of squaring is that the larger numbers carry even more weight than they already do in the least squares fitting procedure. A practical and theoretical investigation into the merits of squaring versus not-squaring has, to our knowledge, not yet been undertaken.

For anova-data a good recommendation is difficult to give. One could model as many means additively as one would deem interesting, and analyse the anova residuals with three-mode principal component analysis. Alternatively, one could set aside only the grand mean, considering the remaining scores as deviations from this grand mean, and analyse them with three-mode analysis, as we did in the *Perceived reality study* in Chapter 7. Different ways of analysing will highlight different aspects of the data, and it is difficult to say beforehand which way is the best.

Kruskal (1981, p.7) criticized subtracting the grand mean, $z_{...}$, from all data points on the grounds that a simultaneous least squares estimation of $z_{...}$ and the components in a three-mode model does not yield the same solution as first estimating the grand mean by least squares, and subsequently the three-mode model. On the basis of this observation he objects to subtracting only the grand mean. His objection could be met as suggested by De Leeuw (1982, pers. comm.), by adding an extra phase to the TUCKALS algorithm in which the grand mean is estimated. Admittedly this was not done in our example in section 7.5.

For some data, such as scores on bipolar scales, considerations connected with substantive theory may suggest choosing the scale midpoints as a 'neutral' zero point, e.g. the midpoint of the scales in semantic differential research (see section 9.4). Deviations of the true means from this neutral point have substantive interest, as are their relationships with the concepts. In the adjective set of the *Cola study* (section 11.2) we also chose this approach.

An objection against removing means, unrelated to the issues discussed above, is their sensitivity, and of least squares estimates in general, to outliers. Such outliers may so badly bias the mean vectors that the transformed data values will be severely biased as well, and their further analysis might not produce the intended results (see section 7.2, and Gnanadesikan & Kettenring, 1972, p.107). The solution in such cases is to deal with the outliers in an appropriate way by using robust measures of centrality like medians. As multivariate outliers (see also Chapter 7), are difficult to detect, spotting them before or after removing means can be a difficult task.

Examples of the suggested centring procedures can be found in the applied chapters of this book, while some reasoned centrings (sometimes combined with standardization) can also be found in the literature, e.g. Gabrielsson (1979, p.162: j-centring, and i-centring; full details in Gabrielsson, & Sjögren, 1974, p.9-11), Hohn (1979, p.167; j-normalization), Gräser (1977, p.83-87; ij-normalization, and j-normalization)).

Finally, as different centrings lead to different solutions, it is preferable to determine a priori which centring is appropriate. Although solutions based on different centring procedures can be inspected to decide a posteriori which scaling is appropriate, one can easily lose sight of the difference between a priori hypotheses and a posteriori conclusions. In the end it will be considerations of research design and subject matter which should decide the appropriate scaling, but the choice is never an easy or automatic one.

6.6 INPUT STANDARDIZATION: COMPARABLE VARIANCES

In comparison with input centring our discussion of standardization will be rather brief, primarily because less research has been done in this aspect of scaling, and because standardization is more complex, and thus less well understood.

A fundamental difference between centring and standardization is that combinations of different centrings do not influence each other, while standardizations do. For instance, double-centring (jk,ik-centring) is a combination of jk-centring and ik-centring. They can be done separately and in any order. For standardization the situation is, however, far more complex: standardization of one mode will generally destroy that of another mode. Iterative procedures have been devised to arrive at stable standardizations for two modes, but the final solution depends on the mode one starts with (see Cattell, 1966a, p.118, and earlier references cited by him). Harshman has provided iterative standardization procedures in his three-mode factor analysis program PARAFAC1 (vide, Harshman & DeSarbo, Note 2, and Kruskal, 1981), but as far as we know technical details are not yet available in print. It seems that for three-mode data conditions can be formulated for unique solutions, but these are not known to us.

The question of iterative standardization bears some resemblance to the problem of iterative proportional fitting for contingency tables (e.g. Bishop, Fienberg, & Holland, 1975). After convergence of the algorithm to perform the proportional fitting the properties of the residuals are known, while for iterative standardization the properties of the double-standardized data are not.

Further complications arise as standardization over one mode may destroy the centring over another. This means that when one wants to have, for instance, a centring over one mode, a normalization of another, and a standardization of a third as Harshman & DeSarbo [Note 2] do in an example, the centring has to become part of the iterative procedure. Harshman & DeSarbo report convergence for the procedure, but it is uncertain what the relationships are between the results of such a procedure and the raw data, the more so because in certain circumstances the order in which the various scalings are performed also seems to have an effect on the solution.

In view of the very incomplete state of affairs in this respect, it is difficult to recommend other standardizations than those in accordance with the centring used, and not requiring an iterative procedure. Considering the definition of the variance, it seems advisable to perform centring first and standardization next, when they are used in conjunction.

In pca-data with variables in the j-mode, standardization will almost always be used together with centring, so as to achieve normalization. This, of course, is the standard practice in standard factor analysis and principal component analysis. With three-mode data, the question remains whether one wants to j-standardize or jk-standardize. Kruskal (1981, p.17, 18) favours j-standardization, because it does not destroy agreement with the (PARAFAC1) model, an argument we discussed above.

An argument put forward by some authors (e.g. Lohmöller, 1979, 1981a; Rowe, 1979) in favour of jk-normalization is that it makes

$$r_{jk,j'k'} = \frac{1}{\ell}\sum_{i=1}^{\ell} \tilde{z}_{ijk}\tilde{z}_{ij'k'} \qquad (j,j'=m;k,k'=1,\ldots,n)$$

$$\tilde{r}_{jj'} = \frac{1}{n\ell}\sum_{k=1}^{n}\sum_{i=1}^{\ell} \tilde{z}_{ijk}\tilde{z}_{ij'k} \qquad (j,j'=1,\ldots,m)$$

$$\ddot{r}_{kk'} = \frac{1}{\ell m}\sum_{i=1}^{\ell}\sum_{j=1}^{m} \tilde{z}_{ijk}\tilde{z}_{ijk'} \qquad (k,k'=1,\ldots,n)$$

correlations, with $\tilde{z}_{ijk}$ the jk-normalized quantities. It is argued that this normalization is advantageous because it allows for "the usual interpretation of the loadings" (Lohmöller, 1979, p. 158). The statement that $\tilde{r}_{jj'}$ and $\tilde{r}_{kk'}$ are correlations is, however, incorrect, as they are only averages of correlations, e.g.

$$r_{jj'} = \frac{1}{n} \sum_{k=1}^{n} \frac{1}{\ell} \sum_{i=1}^{\ell} [(z_{ijk} - z_{.jk})/s_{.jk} \times (z_{ij'k} - z_{.j'k})/s_{.j'k}]$$

$$= \frac{1}{n} \sum_{k=1}^{n} r_{jk,j'k} \quad ,$$

and averages of correlations are not necessary correlations themselves. This is only the case when the $s_{.jk}$ are equal for each j across all k. With centring these problems do not occur as sums of covariances are again covariances, and can be interpreted as 'within sums of squares'.

In mds-data, irrespective of the often recommendable jk,ik-centring, it is at times desirable to standardize matrix-conditionally (k-standardization) for instance, in order to eliminate response styles. The k-standardization can be done without influencing the results from the jk,ik-centring. The reason for this is that both centring and standardization are performed in the same matrix, and centring within a matrix is not influenced if every value in that matrix is divided by a constant.

Our experience with anova-data is very limited, and it is difficult to make a well-founded statement about them. What can be said is that if one takes the anova character seriously, i.e. if the data form a homogeneous set which is assumed to be a very good indicator of one single variable, then overall standardization would be called for, but as mentioned before such standardization does not influence the outcome of the estimation of the parameters in the model.

6.7 INTERPRETATION: GENERAL ISSUES

Whereas the previous sections dealt with preprocessing of raw data, the following sections deal with postprocessing of 'raw' output and its interpretation. Proper handling of input and interpretation of output are always intimately linked with subject matter, and as such difficult to treat generally. What is useful, sensible and clarifying in one case obscures matters in another case. Nevertheless we will try to remain at a level which is as general as possible, and primarily discuss the structural aspects of interpreting the output from three-mode principal component analysis.

One restriction is built into the discussion from the beginning: the fact that it is based on the output from the implementation of the TUCKALS2 and TUCKALS3 algorithms by Kroonenberg (1982a,c). Wherever possible we will make statements with wider implications than only these implementations.

The matrices of component loadings, G, H, and E, together with the core matrix C from the Tucker3 (T3) model,

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

form the basic output from a three-mode principal component analysis with three reduced modes, and G, H, and the extended core matrix $\tilde{C}$ from the Tucker2 (T2) model,

$$z_{ijk} = \sum_{p=1}^{s} \sum_{q=1}^{t} g_{ip} h_{jq} \tilde{c}_{pqk},$$

the basic output from a three-mode analysis with two reduced modes. Some of the characteristics of this output are the following:

- Principal components are columns of orthonormal matrices (G,H,E), i.e. they have length 1, and the scalar products of components within a mode are 0.
- Component matrices are eigenvector matrices of the cross products of the data reduced by the components of the other modes, i.e. of P, Q, and R for the 1st, 2nd, and 3rd mode respectively (see *Theorem 4.1* in section 4.4 for precise definitions).
- Components of a mode are arranged in decreasing order of importance, as expressed by the eigenvalues.
- Eigenvalues or components weights corresponding to the eigenvectors indicate the contribution of the eigenvector or principal component to the overall fit, as expressed by the

$$SS(Fit) = \sum_{i=1}^{\ell} \sum_{j=1}^{m} \sum_{k=1}^{n} \hat{z}^2_{ijk} \text{ , where}$$

$$\hat{z}_{ijk} = \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} \hat{g}_{ip}\hat{h}_{jq}\hat{e}_{kr}\hat{c}_{pqr} \quad (T3), \text{ or}$$

$$\hat{z}_{ijk} = \sum_{p=1}^{s} \sum_{q=1}^{t} \hat{g}_{ip}\hat{h}_{jq}\hat{\tilde{c}}_{pqk} \quad (T2),$$

are the estimated data values based on the fitted model, and

$$\sum_{p=1}^{s} \lambda^{\times}_{p} = \sum_{q=1}^{t} \mu^{\times}_{q} \sum_{r=1}^{u} \nu^{\times}_{r} = SS(Fit)$$

with $\lambda^{\times}_{p}$, $\mu^{\times}_{q}$, and $\nu^{\times}_{r}$ the eigenvalues.

- Standardized eigenvalues or standardized component weights, $\lambda_p(p=1,\ldots,s)$, $\mu_q(q=1,\ldots,t)$, $\nu_r(r=1,\ldots,u)$, are the eigenvalues divided by the total variation in the data expressed by

  $$SS(Total) = \sum_{i=1}^{\ell} \sum_{j=1}^{m} \sum_{k=1}^{n} z^2_{ijk}.$$

- Core matrices are scaled such that

  $$\sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} c^2_{pqr} = SS(Fit) \quad (T3), \qquad \text{or}$$

  $$\sum_{k=1}^{n} \sum_{p=1}^{s} \sum_{q=1}^{t} \tilde{c}^2_{pqk} = SS(Fit) \quad (T2);$$

  this scaling is in accordance with the orthonormality of the components.

In some important aspects these characteristics differ from those of the common representation of the output from Tucker's (1966a) methods.

- Components in Tucker's methods are derived from cross products of raw data, i.e. without taking into account the reduction over the other modes.

- Components in Tucker's methods have been scaled, often by other authors than Tucker, to the size of the eigenvalues.
- Eigenvalues have often been scaled so that the sum over all eigenvalues of a mode is equal to the number of elements in that mode.
- In Tucker's methods the sum of the eigenvalues of different modes generally add up to different values, and the core matrix does not allow an interpretation in terms of sums of squares, unless all components are included.

A discussion of similarities and differences between the ALS approach used here and Tucker's can be found in section 3.6 and Chapter 4.

6.8 INTERPRETATION: COMPONENTS

Components as latent elements. The most common way of interpreting principal components is as latent variables, or in our case also as idealized subjects or prototype conditions. In practical applications these interpretations are often given to all extracted principal components, implying mostly that they represent theoretical constructs in some substantive context. Bargmann (1969), however, rejects the notion that a case can be made in favour of such an interpretation for anything but the first component, because the other components only define directions in the variable space orthogonally to the first one, and this is not necessarily the same as a theoretical construct from some substantive theory.

There are two ways around this - to our opinion correct - statement. One way is to define latent variables, idealized subjects, and prototype conditions, etc. to be the principal components, and to consider the labels attached to these components as convenient summaries of the elements with high loadings on the components without assuming that they necessarily correspond to theoretical constructs. The other way is to assert that the extracted components together define what might be called the 'latent space', which contains the only relevant systematic variation. Then the theoretical constructs which are held responsible for the loadings of the elements on the components, span the latent space. The spatial arrangement of the elements in the latent space is then the representation of the theoretical constructs. Whether one then uses the components or any other set of vectors to interpret the relationships between the variables is immaterial as long as the spatial arrangement is adequately described.

Scaling to the sizes of component weights. As mentioned in section 6.7, in our basic representation of the model the components are scaled to unit length. Adjusting the components in such a way that their lengths are proportionate to their (standardized) weights has certain advantages for plotting the components against one another. Especially when the weights associated with the components are very different, directly plotting them without adjustment might give a wrong impression of their relative importance.

Scaling according to Bartussek. Bartussek (1973) suggested scaling the orthonormal components of a three-mode principal component analysis analoguously to the procedure often encountered in standard principal component analysis. Primarily this means analysing the average cross-product matrices rather than the unaveraged ones, and multiplying the component loadings by the square root of the eigenvalues, as suggested for making plots in the previous paragraph. There it was purely a matter of convenience for plotting, here it is an integrated part of the representation of the model. In order to keep the model essentially the same, the core matrix has to be adjusted with the inverse transformations of those components. The effects of this on the interpretation of the core matrix will be discussed in the next session. In Table 6.3 we have summarized the proposal of Bartussek (1973).

Rotation of components. In standard principal component analysis it is customary to rotate the solution of the variables to some kind of 'simple structure', most often using Kaiser's (1958) varimax procedure. This and other rotational procedures have been extensively applied in three-mode principal component analysis. We will touch upon the rotation issue only very lightly as we have

Table 6.3 *Scaling of output according to Bartussek*

	Tucker	Bartussek
component weights or eigenvalues	$\lambda^\times_p$	$\lambda^*_p = \lambda^\times_p/mn$
	$\mu^\times_q$	$\mu^*_q = \mu^\times_q/\ell n$
	$\nu^\times_r$	$\nu^*_r = \nu^\times_r/\ell m$
components	g_{ip}	$g^*_{ip} = g_{ip}\surd\lambda^*_p = g_{ip}\surd(\lambda^\times_p/mn)$
	h_{jq}	$h^*_{jq} = h_{jq}\surd\mu^*_q = h_{jq}\surd(\mu^\times_q/\ell n)$
	e_{kr}	$e^*_{kr} = e_{kr}\surd\nu^*_r = e_{kr}\surd(\nu^\times_r/\ell m)$
core matrix	c_{pqr}	$c^*_{pqr} = c_{pqr}/\surd(\lambda^*_p\mu^*_q\nu^*_r) = \ell mn\times c_{pqr}/\surd\lambda^\times_p\mu^\times_q\nu^\times_r$

Note: g_p, h_q, and e_r are orthonormal components;
$\lambda^\times_p$, $\mu^\times_q$, and $\nu^\times_r$ are unstandardized eigenvalues

little to add to the standard practice of rotating separate component matrices.

Various authors have advocated particular rotations of component matrices for specific types of data. Lohmöller, for instance, (1981a) recommends rotation of time-mode component matrices to a matrix of orthogonal polynomials as a target, a proposal also put forward by Van de Geer (1974) - see section 13.3. Subject components are often rotated in such a way that the axes pass through centroids of clusters of individuals. Tucker (1972, p.10-12) advocated that the "first priority for these transformations should be given to establishing meaningful dimensions for the object space [of variables]".

The emphasis in the literature on first rotating the component matrices is a consequence of the familiarity with such procedures in standard principal component analysis. In three-mode analysis the core matrix is most difficult to interpret. This suggests concentrating on simplicity of the core matrix rather than on that of the component matrices. By simplicity is here meant a large number of zeroes or very small values in the core matrix, preferably in the off-diagonal elements. Transformations of core matrices to a simple structure were discussed in detail in Chapter 5.

6.9 INTERPRETATION: CORE MATRICES

In this section we will discuss several ways in which the elements of the core matrices of the Tucker3 (T3) and Tucker2 (T2) models can be intepreted. There seem to be at least five, not unrelated ways to do this: (1) percentages of explained variation, (2) three-mode interaction measures, (3) scores of idealized or latent elements, (4) direction cosines, and (5) latent covariations. The last interpretation is far from completely developed and its discussion is deferred until section 13.3.

Explained variation. The core matrix indicates how the components of the three modes relate to one another. For instance, the element c_{111} of the T3 core matrix (Table 6.4) indicates the strength of the relationship between the first components of the three modes, and c_{221} the strength of the relationship between the second components of the first and the second mode in combination

Table 6.4 *Notation for T3 core matrices*

		third mode components					
		1		2		3	
second mode components		1	2	1	2	1	2
first mode components	1	c_{111}	c_{121}	c_{112}	c_{122}	c_{113}	c_{123}
	2	c_{211}	c_{221}	c_{212}	c_{222}	c_{213}	c_{223}
	3	c_{311}	c_{321}	c_{312}	c_{322}	c_{313}	c_{323}

Note: each rectangle is a *frontal plane* of the core matrix with the first of the third mode. The interpretation of the elements of the core matrix is facilitated if one realizes that

$$\sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} c^2_{pqr} = \text{SS(Fit) (T3), and} \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{k=1}^{n} \tilde{c}^2_{pqk} = \text{SS(Fit) (T2)}$$

In other words, each of the c^2_{pqr} indicates how much the combination of the p-th component of the first mode, the q-th component of the second mode, and the r-th component of the third mode contributes to the overall fit of the model, or how the total variation -SS(Total)- is accounted for by this particular combination of components. An analoguous interpretation holds for the T2 core matrix, but now a $\tilde{c}^2_{pqk}$ expresses how much of the total variation is explained by the combination of the p-th and q-th components for the k-th subject or condition. Sometimes it is useful to express $\tilde{c}^2_{pqk}$ as proportion $\sum_p \sum_q \tilde{c}^2_{pqk}$, in order to indicate the importance of that particular combination for the k-th condition or subject.
In addition it can be shown that

$$\sum_{p=1}^{s} \sum_{q=1}^{t} c^2_{pqr}/\text{SS(Total)} = \nu_r \quad (r=1,\ldots,u)$$

$$\sum_{q=1}^{t} \sum_{r=1}^{u} c^2_{pqr}/\text{SS(Total)} = \lambda_p \quad (p=1,\ldots,s)$$

$$\sum_{r=1}^{u} \sum_{p=1}^{s} c^2_{pqr}/\text{SS(Total)} = \mu_q \quad (q=1,\ldots,t)$$

with λ_p, μ_q, and ν_r the standardized component weights for the first, second, and third mode respectively. Similar expressions can be derived for the T2 core matrix. Furthermore

$$\sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} c^2_{pqr}/\text{SS(Total)} = \sum_{p=1}^{s} \lambda_p = \sum_{q=1}^{t} \mu_q = \sum_{r=1}^{u} \nu_r$$

$$= \text{SS(Fit)/SS(Tot)} = \text{Rel.SS(Fit)}$$

In this way the core matrix represents a partitioning of the overall fitted sum of squares - SS(Fit) - into small units through which the (possibly) complex relationships between the components can be analysed. In singular value decomposition (see section 2.2) the squares of the singular values (=eigenvalues of standard principal component analysis) partition the fitted variation into parts which can be attributed to each component. The core matrix does the same for the combination of three components, and in this sense three-mode principal component analysis is for three-mode data the direct analogue of what singular value decomposition is for two-mode data.

Three-mode interactions. In singular value decomposition of two-mode matrices, say, Z = GΛH', the p eigenvectors in G and H are pairwise identical, as can be seen from the sum notation

$$z_{ijk} = \sum_{p=1}^{s} \lambda_p g_{ip} h_{jp}$$

The matrix Λ contains the measures of the interactions between the components of the i-mode G, and those of the j-mode H. For two-mode data $\lambda_{pp'}=0$ if $p \neq p'$; so there is no interaction between the vectors g_p and $h_{p'}$, but only between g_p and h_p, which thus have an exclusive interaction with one another. In the two-mode case it is, therefore, legitimate and customary to think of just one set of components for which we have loadings for the j-mode elements H, and scores for the i-mode elements, GΛ (see also section 2.2, Figure 2.1). The strength of the interaction is usually expressed as λ^2_{pp}, or amount of variation accounted for, as discussed above.

The interaction structure between the components of three modes can be, and usually is far more complex. The parallel structure to the two-mode situation would be a body-diagonal (s×s×s) core matrix, so that g_p, h_q, and e_r would only have a non-zero interaction if p=q=r. This is in fact the situation postulated for the PARAFAC1/CANDECOMP model

$$z_{ijk} = \sum_{p=1}^{s} c_{ppp} g_{ip} h_{jp} e_{kp},$$

A core matrix with only non-zero body-diagonal elements has the most 'simple' structure a core matrix can have, and the interpretation can be relatively straightforward - from a technical point of view.

In a s×s×u core matrix a still relatively 'simple structure' is $c_{pqr} \neq 0$ if p=q, r=1,...,u, and zero elsewhere. In this case all frontal planes, C_r, of the core matrix are diagonal (for an example see section 10.5, Table 10.2). When u=n, we have a T2 diagonal extended core matrix with diagonal frontal planes, C_k (for an example see section 2.8, Table 2.7).

A similar simple structure is sometimes found for r=2:

$c_{pqr} \neq 0$ for p=q, r=1; $c_{pqr} \neq 0$ for q=s-p+1, r=2;

and zero elsewhere.

Here the first frontal plane is diagonal, and the second is 'anti-diagonal', i.e. the diagonal running from the bottom lefthand corner to the upper righthand corner is non-zero. For an example of such a core matrix see section 15.8, Table 15.5.

The structure in a three-mode core matrix is often not simple, and thus interpretational complexities arise, as a component will have interactions with more than one component of another mode. One of the complications is due to the interpretation of the sign of a core element and the fact that the interactions refer to continuous rather than discrete entities, unlike interactions between levels of factors in analysis of variance, and categories in contingency tables.

Suppose that c_{pqr} has a positive sign, so that the interaction of the p-th, q-th, and r-th component of the first, second, and third modes respectively is positive. This positive interaction indicates that four different combinations of elements of the three modes tend to occur together in the data:

(+,+,+); (+,-,-,); (-,+,-); (-,-,+),

in which a plus (minus) on the p-th, q-th, and r-th place in (p,q,r) refers to positive (negative) loadings on the p-th component of the first mode, q-th component of the second mode, and the r-th component of the third mode, respectively. A parallel formulation can be that "positive loadings on p occur together with loadings of the same sign of q and r: (+,+,+) and (+,-,-) negative loadings on p occur with loadings of opposite signs on q and r: (-,+,-) and (-,-,+). A negative sign of c_{pqr} corresponds with the joint occurrence of the combinations (+,-,+), (+,+,-), (-,+,+), and (-,-,-). The mental juggling with combinations of positive and negative loadings of different components is what makes the interpretation very difficult in many cases. In some data sets certain components have only positive loadings which simplifies the interpretation, as the number of combinations reduces with a factor two. Sometimes certain core elements are so small that they need not be interpreted, which also simplifies interpretation. In section 8.4 we have given a detailed analysis of a complex core matrix as an example of how to deal with the problem of interpreting the combinations of negative and positive loadings on different components.

A good strategy to simplify the interpretation is to make 'conditional' statements by only making statements about elements which have, for instance, positive loadings on a component. The core matrix then represents only the interaction between the loadings of the two other modes, 'given' the positive loading on the third. The joint plots and component scores discussed in section 6.10 are examples of such an approach. In practice we have observed that it is most useful to use the third mode (subjects, conditions) for 'conditioning'. To carry the argument a bit further, one might say that the T2 extended core matrix is also an example of conditioning as no components of the third mode exist, and one can deal with the interactions between the components of the first two modes one frontal plane (= one element of the third mode) at a time.

Scores of idealized elements. This interpretation was the basis for the second explanation of the three-mode principal component model in section 2.2, and is the one usually employed in the literature. Each element of the core matrix is assumed to represent the score of an 'idealized subject' on a latent variable in a prototype condition. The main difference with the interpretation in the previous subsection is that there the interpretation was based on interactions between loadings on components, while here we construct interpretations in terms of the components themselves.

It depends on the applications which way will be easier to handle. When one rejects the idea of labelling components the latter method is in fact not applicable. In examples with very few variables and conditions the labelling of components is in any case a rather risky business, and the former approach seems more helpful. In other applications, especially when the labelling of the components is firmly established, the latter approach might be easier to use.

Part of the purpose of Bartussek's (1973) proposal to scale the components (see section 6.8) was to allow an interpretation of the core matrix as scores of idealized quantities, as in the paragraph above, but with the specific characteristic that the absolute size of the elements in the core matrix is independent of the amount of variation accounted for by the components. In this way an element of the core matrix c^*_{pqr} (see Table 6.3), reflects a score of a real subject who has a loading of *one* on the p-th subject component and zero on the others for that real variable which loads *one* on the q-th component and zero elsewhere, in the condition which loads *one* on the r-th component and zero elsewhere. In this way the elements of the core matrix reflect a very 'pure' subject, variable and condition (see Bartussek, 1973, p.179).

Notwithstanding the correctness of Bartussek's interpretation of his scaled core elements, it is doubtful whether his scaling procedure is really necessary. His argument centres around the independence of the core elements from the amount or variation accounted for by the respective components to which they refer. It is, however, exactly this dependence on the variation accounted for which makes it possible to assess the relative importance of these core elements. It is, by the way, a misnomer to call the adjusted core elements, c^*_{pqr} 'factor scores' of idealized subjects, variables, etc., as Bartussek does. Just 'scores' is more appropriate, as the core elements are the scores of the idealized subjects on latent variables in prototype conditions, so that the word 'factor' or 'component' confuses the issue - the more so because in three-mode analysis already a definition of 'component scores' exists (see section 6.10).

Two more points should be mentioned with respect to Bartussek's (1973) proposals. The 'real person'-'real variable'-'real condition' combination with a loading *one* for just one of the components of each mode might easily be a non-existent or impossible combination in a particular data set, i.e. such a point might lie far away from all other points in any one or all of the component spaces. Cliff (1968) and Ross (1966) raised objections against the use of such idealized quantities as was proposed by Tucker & Messick (1963), see also Tucker (1972, p.26).

The second point is that the scaling can lead to some absurdly large values for those core elements associated with very small eigenvalues (see Table 6.3).

Direction cosines. In those cases where two modes are equal or the components define the same space, an additional interpretation of the core matrix is possible. Within the context of multidimensional scaling of individual differences, for instance, the input similarity matrices satisfy these conditions, and within this field an interpretation has been developed in terms of correlations and direction cosines of the axes of the spaces common to two (generally the first and second) modes (see section 3.2, Tucker, 1971, p.7, and Carroll & Wish, 1974, p.91).

In these situations it makes sense to speak about the angle between the first and second component of the common space. This angle can be derived from the off-diagonal elements of the core planes, as they can be looked upon as a direction cosine or correlation between component p and component q, provided $\tilde{c}_{pqk}$ is scaled by dividing it by $\tilde{c}^{\frac{1}{2}}_{ppk}$ and $\tilde{c}^{\frac{1}{2}}_{qqk}$, and the components are standardized. The direction cosine indicates the angle under which the k-th condition 'sees' the axes or components of the common space (for an example see section 2.8).

This approach towards the core matrices follows Tucker's three-mode scaling (1972) and Harshman's PARAFAC2 (1972) as pointed out, for instance, in Carroll & Wish (1974) and Dunn & Harshman (1982). The joint plots, to be discussed in the next section, are more in line with Carroll & Chang's approach to treating the core matrix (see references above), in which the (extended) core matrix is decomposed by either eigenvalue-eigenvector or singular value decompositions.

6.10 INTERPRETATION: JOINT PLOTS AND COMPONENT SCORES

Various kinds of auxiliary information can be useful for the interpretation of results from a three-mode principal component analysis. In this section we will discuss the what we will call *joint plots*, which display the elements of different modes in the same figure, and *component scores*, which are the scores of, for instance, each subject-condition combination on the components of the variables.

Joint plots. After the components have been computed, the core matrix will provide the information about the relationships between these components as discussed in the previous section. It is very instructive to investigate the component loadings of the subjects jointly with the component loadings of, for instance, the variables, by projecting them together in one space, as it then becomes possible to display the interaction between variables and subjects. The plot of the common space will be called a *joint plot.*

Such a joint plot of every pair of component matrices, say G and H, for each component of the third mode, say E, in the Tucker3 case, and for the average core plane in the Tucker2 case, is constructed in such a way that g_p (p=1,...,s) and h_q (q=1,...,t) -i.e. the columns of G and H respectively - are close to each other. Closeness is measured as the sum of all s×t squared distances $d^2(g_p,h_q)$ over all p and q.

The plots are constructed as follows. For each component r of E, the components G and H are scaled by dividing the core plane associated with that component, C_r, between them (using singular value decomposition), and weighting the scaled G and H by the relative number of elements in the modes to make the distances comparable:

$$D_r = GC_rH' = G(U_r\Lambda_rV_r')H' = (\frac{\ell}{m})^{\frac{1}{4}} (GU_r\Lambda_r^{\frac{1}{2}}) (\frac{m}{\ell})^{\frac{1}{4}} \overset{\sim\sim}{(HV_r}\Lambda_r^{\frac{1}{2}})' = \tilde{G}_r\tilde{H}_r'$$

with

$$\tilde{G}_r = (\frac{\ell}{m})^{\frac{1}{4}} GU_r\Lambda_r^{\frac{1}{2}} \text{ and } \tilde{H}_r = (\frac{m}{\ell})^{\frac{1}{4}} HV_r\Lambda_r^{\frac{1}{2}} \quad r=1,...,u.$$

For the rationale behind this construction, and a more detailed discussion, see Kroonenberg & De Leeuw (1977). When C_r is not square only the first a = min(s,t) components can be used. The procedure can be interpreted as rotating the component matrices by an orthonormal matrix, followed by a stretching (or shrinking) of the rotated components. Similar procedures for plotting two sets of vectors into one figure have been developed by Schiffman & Falkenberg (1968; see also Schiffman et al., 1981, Ch. 14), Gabriel (1971; biplot), Carroll (1972; MDPREF), Benzécri (1973; correspondence analysis; see also Gifi, 1981, Ch. 4).

In practice joint plots have proved a powerful tool for disentangling complex relationships between components, and nearly every example in this book uses joint plots in one way or another. If we designate the first mode element as subjects and the second mode elements as variables, then we may say that both the subjects and variables in a joint plot are vectors from the origin. Their distances, d^r_{ij}, are the scalar product between the vectors, i.e.

$$d^r_{ij} = \sum_{p=1}^{a} \tilde{g}_{ip}\tilde{h}_{jp}, \text{ with } a = \min(s,t).$$

By projecting, for instance, subjects on a particular variable the relative importance of that variable for the subjects can be assessed from the size and the sign of the resultants. One of the advantages of the joint plot is that the interpretation of the relationships of the variables and the subjects can be made directly without involving components or their labels, given the r-th component for conditions. Another feature of the joint plots is that via the core plane C_r the axes of the joint plot are scaled according to their relative importance, so that visually one obtains a correct impression of the spread of the components.

Component scores. In some applications it is useful to inspect the scores of all combinations of the elements of two modes on the components of the third mode. For instance, for longitudinal data the scores of each subject-time combination (or ik-combination) on the variable (j) components can be used to inspect the development of an individual's score on the latent variable over time. In some examples, these component scores in fact turn out to be a very succesful summary of the relationships involved (see Chapters 2 and 8). They serve as an intermediate level of condensation between the raw data, and the three-mode model.

The component scores on the r-th component of the third mode have the form

$$d_{ijr} = \sum_{p=1}^{s} \sum_{q=1}^{t} g_{ip}h_{jq}c_{pqr} \text{ or } D_r = GC_rH' ,$$

but by using other combinations of component matrices, three different sets of scores can be calculated. In general, only a few of these will be useful in a particular application.

One of the interesting aspects of the component scores d_{ijr} is that they are at the same time the inner products

$$\sum_{p=1}^{a} \tilde{g}^r_{ip}\tilde{h}^r_{jp} ,$$

thus expressing the closeness of the elements from different modes in a joint plot.

Sometimes it is not useful to display the component scores for different components in one plot, but it is clearer to plot the component scores of, for instance, the subjects for each of the conditions (see section 2.10, Fig. 2.7; and section 8.5, Fig. 8.4). Such plots are sometimes easier to use, explain, or present than joint plots in which one has to inspect projections on vectors.

In case of a good approximation of the model to the data the component scores as described above will resemble the component scores from a standard principal component analysis on a data matrix in which the columns are variables, and the rows the subject-condition combinations. Other writers too, have suggested using such component scores (e.g. Hohn, 1979).

Mixed-mode matrices. A somewhat different, to our mind incorrect, approach towards inspecting measures of elements of one mode on the component of another was taken by Wainer, Gruvaeus, & Zill (1973). They defined what they called *mixed-mode matrices*. If we designate M(e,f) as the mixed-mode matrix of 'loadings' (as Wainer et al, called them) of the elements of the e-th mode on the components of the f-th mode, and choose e=1, and f=2 for illustration sake, then $M(1,2) = \{m^{12}_{iq}\}$, or $\{m_{iq}\}$ for short, is defined as

$$m_{iq} = \sum_{k=1}^{n} \sum_{j=1}^{m} \sum_{p=1}^{s} \sum_{r=1}^{u} g_{ip} h_{jq} e_{kr} c_{pqr} = \sum_{k=1}^{n} \sum_{j=1}^{m} h_{jq} d_{iqk} ,$$

with $d_{iqk} = \sum_{p=1}^{s} \sum_{r=1}^{n} g_{ip} e_{kr} c_{pqr}$.

The d_{iqk} are the component scores of the subject-condition combinations on the variable components, defined in the previous paragraph. We may rewrite the m_{iq} further as

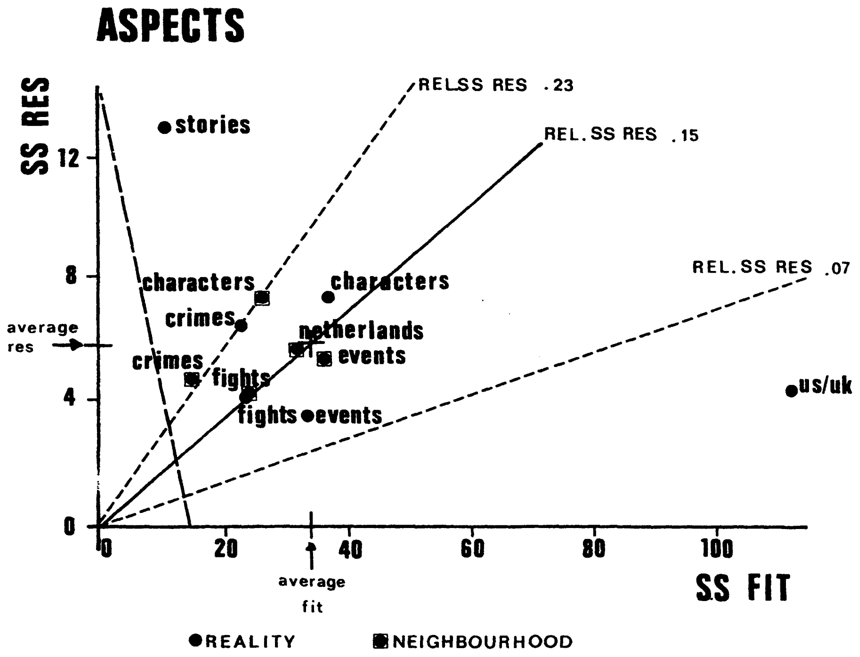
$$m_{iq} = n \sum_{j=1}^{m} h_{jp} \{\frac{1}{n} \sum_{k=1}^{n} d_{iqk}\} = n \sum_{j=1}^{m} h_{jq} \bar{d}_{iq.} = n\bar{d}_{iq.} \{ \sum_{j=1}^{m} h_{jq} \} =$$

$$= nm\bar{d}_{iq.}\ \bar{h}_{.q} .$$

The righthand side is the product of the component score of subject i averaged over conditions, and multiplied by the average loading on variable component q. Thus the m_{iq} are weighted average component scores. One of the problems with mixed-mode matrices is the average component loading $h_{.q}$. For centred data entire mixed-mode matrices may become zero, because all $h_{.q}$, and/or analoguous component averages can be zero. In view of this, it seems better if one wants to have something like mixed-mode matrices to use the average component score directly.

RESIDUALS

7


ASPECTS
SS RES
12
8
4
0
average res
●stories
characters
crimes
crimes
fights
fights
characters
netherlands
events
●events
●us/uk
RELSS RES .23
REL.SS RES .15
REL.SS RES .07
0
20
40
60
80
100
average fit
SS FIT
●REALITY
▨NEIGHBOURHOOD

7.1 INTRODUCTION

Until now, three-mode principal component analysis has been, both in this book and elsewhere, primarily used for *summarization*, i.e. describing a large body of data with a small(er) number of more basic statistics. As Gnanadesikan & Kettenring (1972, p.81) state, it would be very desirable if a method would also be useful for *exposure*, i.e. not only detecting the anticipated, but also unanticipated characteristics of the data. In fact, they argue "from a data-analytic standpoint the goals of summarization and exposure ought to be viewed as two sides of the same coin". For general discussions on the role and use of summarization and exposure in data analysis see Tukey & Wilk (1966) and Gnanadesikan & Wilk (1969).

In this chapter we will discuss exposure via the analysis of residuals from a three-mode principal component analysis. In section 4.3 the basic equations for separating the fit of the model (summarization) and the residuals (exposure) were explained. Here we will first discuss the analysis of residuals in a more general (theoretical) framework. Then we will review existing proposals for two-mode analysis, and comment on their three-mode generalizations. Finally we will illustrate and evaluate the proposals with one particular data set.

7.2 INFORMAL INFERENCE AND GOALS OF RESIDUAL ANALYSIS

In the analysis of interdependence (see also section 6.4), such as, for instance principal component analysis and multidimensional scaling, the fit of the model is generally inspected, but no standard practice exists of examining the residuals as is usual in the analysis of dependence, especially in regression analysis. Loglinear analysis of multi-way tables forms a welcome exception. It should be realized that in three-mode analysis there may be a good fit of the model to the data, but a distorted configuration in the projected space (e.g. the space defined by the first principal components) due to some isolated points far away from the best fitting space. The examination of such isolated points is of interest in practical work, and may provide an interpretation of the nature of the heterogeneity in the data (cf. Rao, 1964, p.334).

"One of the most insightful processes for exposure [of such isolated points] in analysis of uniresponse data is the study of residuals in a variety of ways [....]. The development of [....] methods for analyzing residuals and for detecting outliers for multivariate situations is very difficult, [....]. A general point is that multiresponse models or modes of summarization are inherently more complex and ways in which the data can depart from the model are infinitely more varied than in the univariate case. Consequently, it is all the more essential to have informal, informative summarization and exposure procedures" (Gnanadesikan & Kettenring, 1972, p.82).

A similar emphasis on using informal techniques for looking at residuals rather than formal statistical procedures can be found in Tukey (1968, p.274): "Since residuals are mainly used to look for what might be going on beyond what is already in the model, we can go for this with first-order answers. Small corrections whether or not of higher order, are often negligible [....]. The important thing is to look at 'residuals'; details of definition matter less".

An informal analysis of residuals of three-mode analysis is, however, not without hazards. The specific structure of the data, i.e. the three-way design and the initial scaling may introduce constraints on residuals or subsets of residuals. Furthermore, the presence of outliers, in particular outlier interactions among the three modes, may affect more than one summary measure of the residuals, and thereby distort conclusions drawn from them. In short, all the woes of the regular analysis also pertain to the analysis of residuals, which is performed to detect inadequacies in the former (see also Barnett & Lewis, 1978, Ch. 6).

There seem to be three major *goals* for which residuals are used, and which are relevant for three-mode analysis:

(i) detection of *outliers*, i.e. points which appear to deviate relative to the chosen model from the other members of the sample;

(ii) detection of *influential points*, i.e. points which determine for a large part the solution of the model;

(iii) detection of *unmodelled systematic trends*, i.e. trends which are not (yet) fitted by the model, because not enough components have been included, or because they are not in accordance with the model itself.

7.3 PROCEDURES FOR ANALYSING RESIDUALS

In the major exposition on the treatment of multidimensional residuals (Gnanadesikan & Kettenring, 1972; reprinted in slightly revised form as section 6.4. in Gnanadesikan, 1977; summarized in Barnett & Lewis, 1978, Ch. 6, especially section 6.2) a distinction is made between residuals arising from principal component analysis and least squares residuals arising from multivariate linear models. Below we will show that the distinction is not as clear-cut as claimed by Gnanadesikan & Kettenring, and that, therefore, procedures developed for both cases can be useful in (three-mode) principal component analysis.

Principal component residuals from two-mode data. Standard "principal component analysis may be viewed as fitting a set of mutually orthogonal hyperplanes by minimizing the sum of squares of orthogonal deviations of the observations from each plane in turn. At any stage, therefore, one has residuals that are perpendicular deviations of the data from the fitted hyperplane" (Gnanadesikan, 1977, p.259). Thus, given that t principal components are necessary to describe a particular data set adequately, the projections of the m-dimensional data on the last i.e. (m-t) not-fitted principal components will be relevant for assessing the deviation of a subject i from the t-dimensional fitted hyperplane or subspace. Rao (1964, p.334) suggested using the length of the perpendicular of each subject i on the best fitting space for detecting its lack of fit.

If we write the standard principal component model as in Fig. 2.1:

$$z_{ij} = \sum_{q=1}^{t} a_{iq}h_{jq} \qquad (i=1,\ldots,\ell;j=1,\ldots,m), \text{ or } Z=AH'$$

with a_{iq} the component score for the i-th subject on the q-th component, and h_{jq} the loading of the j-th variable on the q-th component, then the component score a_{iq} can be written as

$$a_{iq} = \sum_{j=1}^{m} z_{ij}h_{jq} \qquad (i=1,\ldots,\ell;q=1,\ldots,t).$$

The squared length of the projection of the vector in the variable space representing subject i on the vector in the variable space representing component q is thus a^2_{iq}, and Rao's proposal is to use

$$d^2_i = \sum_{q=t+1}^{m} a^2_{iq} = \sum_{q=t+1}^{m} \{ \sum_{j=1}^{m} z_{ij}h_{jq}\}^2, \text{ or}$$

$$d^2_i = \sum_{j=1}^{m} z^2_{ij} - \sum_{q=1}^{t} \{ \sum_{j=1}^{m} z_{ij}h_{jq}\}^2$$

for assessing the fit of subject i.

It is not difficult to show that the d^2_i are nothing but the sum of squares of the least squares residuals for the i-th subject as, for instance, is shown by Rao (1964). The solution of the standard principal component analysis can be found by the minimization of a least squares loss function. The approximate solution $\tilde{Z}$ with t components follows from the minimization of

$$[[Z-\tilde{Z}]]^2 = \sum_{i=1}^{\ell} \sum_{j=1}^{m} (z_{ij}-\tilde{z}_{ij})^2 = \sum_{i=1}^{\ell} \sum_{j=1}^{m} (z_{ij}- \sum_{q=1}^{t} a_{iq}h_{jq})^2$$

As in section 4.3 we may rewrite the loss function at its minimum in separate sums of squares

$$\sum_{i=1}^{\ell}\sum_{j=1}^{m}(z_{ij}-\tilde{z}_{ij})^2 = \sum_{i=1}^{\ell}\sum_{j=1}^{m} z_{ij}^2 - \sum_{i=1}^{\ell}\sum_{j=1}^{m}\tilde{z}_{ij}^2, \text{ or}$$

SS(Res) = SS(Total) - SS(Fit).

The second term on the right hand side may be rewritten as

$$\sum_{i=1}^{\ell}\sum_{j=1}^{m}\tilde{z}_{ij}^2 = \sum_{i=1}^{\ell}\sum_{j=1}^{m}\{\sum_{q=1}^{t} a_{iq}h_{jq}\}^2 = \sum_{i=1}^{\ell}\sum_{j=1}^{m}\{\sum_{q=1}^{t} h_{jq}[\sum_{j'=1}^{m} z_{ij'}h_{j'q}]\}^2$$

which after some algebraic manipulation using the columnwise orthonormality of H becomes:

$$\sum_{i=1}^{\ell}\sum_{j=1}^{m}\tilde{z}_{ij}^2 = \sum_{i=1}^{\ell}\sum_{q=1}^{t}\{\sum_{j=1}^{m} z_{ij}h_{jq}\}^2 .$$

Thus

$$\sum_{i=1}^{\ell}[\sum_{j=1}^{m}(z_{ij}-\tilde{z}_{ij})^2] = \sum_{i=1}^{\ell}[\sum_{j=1}^{m} z_{ij}^2 - \sum_{q=1}^{t}\{\sum_{j'=1}^{m} z_{ij'}h_{j'q}\}^2].$$

The expression between square brackets is exactly Rao's measure for assessing the fit of each subject. The squares in the expressions ensure that for each subject i

$$d_i^2 = \sum_{j=1}^{m}(z_{ij}-\tilde{z}_{ij})^2 .$$

The importance of the above results is that in standard principal component analysis the difference between the least squares and the principal component residuals is not as clear-cut as suggested by Gnanadesikan & Kettenring (1972). The major difference is that in principal component analysis the predictors are latent variables (see section 6.8 for a discussion of this interpretation of principal components), and in (multivariate) regression the predictors are measured quantities. The similarity of the two kinds of residuals implies that proposals put forward for least squares residuals should also be workable in principal component analysis.

Gnanadesikan & Kettenring (1972, p.98) supplemented Rao's proposal by suggesting a gamma probability plot for the d_i^2 to search for aberrant points. We have, however, no experience with the procedure and do not propose to go into it. They also suggested (p.99) using various plots involving the last few principal components, but as in three-mode analysis in its alternating least squares form these are not available, we will not go into these proposals either.

Jackson (with various co-authors) is one of the few next to Gnanadesikan, Rao, and Hawkins (1974, 1980) dealing explicitly with residuals from principal component analysis (Jackson, 1959,1980; Jackson & Bradley, 1966; Jackson & Morris, 1957; Jackson & Mudholkar, 1979). Jackson & Mudholkar (1979) discuss a series of statistics for testing the overall adequacy of a principal component model. Their main statistic is identical to Rao's Σd_i^2, but they start from the residual sum of squares end, and fail to recognize the identity. In fact, they go to some length to expound the advantages of their own approach over Rao's.

Least squares residuals from two-mode data. As noted above, least squares residuals are generally assumed to have arisen from (multivariate) linear models with measured predictors. Any review of the treatment of such residuals is outside the scope of this book. We will only discuss some proposals which could be of relevance for linear models with latent predictors, such as principal component analysis. Again we will lean heavily on Gnanadesikan & Kettenring (1972) and Gnanadesikan (1977), as they seem to be the only sources for treatment of multivariate or multidimensional residuals in this context. Barnett & Lewis (1978, Ch. 7) review methods, mainly tests, for outlier detection in designed experiments and regression. Some of these might be useful in principal component analysis, but on the whole they are too specific to be readily applied in a first-order investigation of residuals from three-mode principal component analysis, and they will therefore not enter our discussion.

There are two basic ways of looking at individual residuals, $\varepsilon_{ij} = z_{ij} - \hat{z}_{ij}$ $(i=1,\ldots,\ell;j=1,\ldots,m)$, with $\hat{z}_{ij}$ the fitted value on the basis of the model. (Note that in the previous paragraph we looked at summary measures for the residuals of an element of a mode, rather than at each residual ε_{ij} separately.) The first is to treat the residuals as an *unstructured sample*, and employ techniques for investigating such unstructured samples, such as found in Gnanadesikan (1977, p.265):

1. Plotting the residuals against certain external variables, components, time (if relevant), or predicted (=fitted) values ($\hat{z}_{ij}$). Such plots might highlight still existing systematic relationships [Goal (iii)], or show, for instance, an unusually small residual relative to the fitted values, possibly indicating that this residual is associated with an overly dominant point, or combination of points [Goal (ii)].

2. Producing one-dimensional probability plots of the residuals, e.g. full normal plots for the residuals, or χ_1^2 plots for squared residuals. According to Gnanadesikan (1977, p.265) "Residuals from least squares procedures with measured predictors seem to tend to be 'supernormal' or at least more normally distributed than original data". The probability plots may be useful in detecting outliers [Goal (i)], or other peculiarities of the data, such as heteroscedasticity. For a full explanation of probability plotting, see Gnanadesikan (1977, p.197-200), or Wilk & Gnanadesikan (1968).

The second overall approach to residuals is to take advantage of the structured situation from which the residuals arose, i.e. from a design with specific meanings for the rows and columns, in other words to treat them as a *structured sample*. As mentioned in section 7.2, exploiting the specific design properties may introduce problems, exactly because of the design, and because the constraints put on the input data which may influence the residuals.

Three-mode residuals. So far we have only discussed the analysis of two-mode residuals, but it should be clear from the above that a third mode does not introduce many new complications. The unstructured approach remains essentially the same. It might, however, be useful in certain cases to consider several unstructured samples, for instance one for each condition. This might especially be appropriate in the case of so-called matrix-conditional or mds-data (see section 6.2), i.e. data in which the measures for each k-mode element have been generated independently. (For a discussion on matrix conditionality see, for instance, Young, De Leeuw, & Takane, 1980, or Young, 1981.) For the structured approach it seems most useful to look at a three-way partitioning of the residuals into sums of squares for the elements of each mode separately (à la Rao) (see Table 7.5).

With all proposals for carrying out a residual analysis it should be kept in mind that we are interested in rough, or first-order results. Attempting to perfect such analyses by adding more subtlety carry with them the danger of attacking random variation. After all we are only dealing with measures from which, in principle, the main sources of variation have already been removed.

7.4 SCHEME FOR THE ANALYSIS OF THREE-MODE RESIDUALS

In this section we will present an *analysis scheme* for the treatment of the residuals from a three-mode principal component analysis. It is felt that by following this scheme a reasonable insight will be gained into the nature of the residuals, so that decisions can be made about the quality of the three-mode solution obtained, and about the need for further analysis. The use of the scheme will be further explained and illustrated in section 7.6 in connection with the data from section 7.5.

Table 7.1 presents the scheme for the analysis of three-mode residuals. A distinction is made between analysing regular residuals and squared residuals. For the first step the use of squared residuals via Rao's distance function is preferred, because fewer numbers have to be investigated, and use is made of the structure present in the residuals. Moreover, the resulting numbers have a direct interpretation in terms of variation accounted for, and the sums of squares can directly be compared with the overall (average) fitted and residual sums of squares. Finally, any irregularity is

Table 7.1 *Analysis scheme for three-mode residuals*

Given an $\ell \times m \times n$ block of *residuals*, $\varepsilon_{ijk} = z_{ijk} - \hat{z}_{ijk}$, and
an $\ell \times m \times n$ block of *squared residuals*, $\varepsilon^2_{ijk} = (z_{ijk} - \hat{z}_{ijk})^2$

Step 1 *(structured approach)*

Investigate the *squared residuals* per element of a mode, i.e. look at

$$SS(Res_i) = \sum_{j,k} \varepsilon^2_{ijk};\ SS(Res_j) = \sum_{k,i} \varepsilon^2_{ijk};\ SS(Res_k) = \sum_{i,j} \varepsilon^2_{ijk}$$

A. Inspect the *distributions* of the

- SS(Total)s of the elements per mode, to detect elements with very large SS(Total)s which might have a large influence on the overall solution, and elements with a very small SS(Total), which do not play a role in the solution,
- SS(Res)s of the elements per mode to detect ill-fitting and well-fitting points,
- Relative SS(Res)s to detect relative differences between elements; e.g. Rel. $SS(Res_i) = SS(Res_i)/SS(Tot_i)$ $(i=1,\ldots,\ell)$.

by using histograms, stem-and-leaf displays, probability or quantile plots, etc.

B. Use *sums-of-squares plots* for each mode to facilitate the evalutation of well-fitting and ill-fitting points.

C. Use an *analysis-of-variance decomposition* of the squared residuals to compare the residual sums of squares across modes.

Suggested action:

- IF no irregularities AND acceptable fit STOP, OR for surety GOTO *step2*.
- IF no irregularities AND unacceptable fit GOTO *step2* AND/OR increase number of components AND redo the analysis.
- IF one element of each mode (say, $\hat{\imath},\hat{\jmath},\hat{k}$) BOTH fits badly AND has a very large SS(Total), check for clerical error at data point $(\hat{\imath},\hat{\jmath},\hat{k})$.
- IF some elements of any mode fit badly GOTO *step2*.
- IF one element f of a mode has a very large $SS(Total_f)$, AND a very small $SS(Res_f)$, delete this element AND redo the analysis to assess the influence of this element on the overall solution, OR rescale input, especially equalize variation in that mode, AND redo the analysis.

Table 7.1 (cont'd)

Step2 *(unstructured approach)*

Investigate the *residuals* as an unstructured sample

A. Examine the *distribution* of the residuals via a normal probability (or quantile) plot.

B. Examine *plots* of the residuals, ε_{ijk}, versus

- *fitted values*, $\hat{z}_{ijk}$, for trend, systematic patterns or unusual points,
- *data* for remaining trends,
- *external variables* for identification of systematic pattern in the residuals.

Suggested action:

- IF trends, or systematic patterns have been found THEN re-examine the appropriateness of the model AND STOP, OR describe these patterns separately AND STOP, OR increase the number of components AND redo the analysis.
- IF a few large residuals are present AND no systematic pattern is evident THEN check the appropriate data points, AND/OR STOP.
- IF no large residuals or trends are present STOP.

enhanced by the squaring, and no cancellation due to opposite signs occurs during summing. For the second step the regular residuals are preferred, especially because at the individual level the signs are important to assess the direction of the deviation and to discover trends.

The problem of examining the individual residuals can, however, be considerable. In the first place, the number of points to be examined is exactly as large as before the three-mode analysis, i.e. no reduction has taken place in the number of points to look at. Secondly, when attempting to use a technique such as analysis of variance with a standard statistical package which uses a general linear model approach with a design matrix, the storage problems can easily run out of hand as in the MANOVA subprogram of SPSS (Hull & Nie, 1981).

A partial solution might be to examine the residuals matrix-conditional as suggested above.

7.5 AN ILLUSTRATIVE DATA SET: PERCEIVED REALITY STUDY

In this section we describe briefly the three-mode analysis of the data set we use to illustrate the procedures for the analysis of three-mode residuals.

The data were collected from primary school children, to acquire information on the way children perceive reality as shown in TV serials. The children rated 8 film types on 11 three-point rating scales, each of which attempted to measure some aspect of the *perceived reality* of the film types (see Table 7.2). Scores were aggregated per grade, so that averages were available for the grades three, four, five and six. Similar scores were available for students of a teacher training college (an 'adult' control group), so that the complete data set consisted of 11 *aspects* (i-mode), 8 *film types* (j-mode), and 5 *"grades"* (k-mode). Before analysis the overall average (1.67) was subtracted from each data point in order to keep the contribution of the overall average to the total sum of squares outside the analysis (see also section 6.5). The size of the overall average indicates, by the way, that the perceived reality over all film types, aspects, and grades is below the midpoint of the scale (2), and thus on the unrealistic side.

As the aim of the present chapter is to examine residuals rather than give detailed substantive explanations and interpretations, we will bypass the original research questions and their motivations. These can be found in Van der Voort (1982). Here we will mainly give a summary description of the three-mode results.

The way the grades used the relationships between aspects and film types is the primary focus of this analysis. The aspects were divided into two classes called *neighbourhood* or agreement items, and *realism* items (see Table 7.2) on the basis of a standard prin-

Table 7.2 *Perceived reality study: Data description*

*Aspects**

1.	Was the film based on a *real story?*		
2.	Could these *events*	occur in *reality?*	*realism* aspects
3.	Could these *characters*	occur in *reality?*	
4.	Could these *fights*	occur in *reality?*	
5.	Could these *crimes*	occur in *reality?*	
6.	Could these *events*	occur in the *US/UK?*	
7.	Could these *events*	occur in your *neighbourhood?*	*neighbourhood* aspects
8.	Could these *characters*	occur in your *neighbourhood?*	
9.	Could these *fights*	occur in your *neighbourhood?*	
10.	Could these *crimes*	occur in your *neighbourhood?*	
11.	Could these *events*	occur in *The Netherlands*	

Film types

1.	*Cowboy*films	*unrealistic* films
2.	Films like the *Hulk*	
3.	*Cartoons*	
4.	*Chivalry* films	
5.	*Children's adventure* films	*realistic* films
6.	Detective films with *male detectives*	
7.	Detective films with *female detectives*	
8.	*Police* films	

Grades

1. *Third* grade primary school
2. *Fourth* grade primary school
3. *Fifth* grade primary school
4. *Sixth* grade primary school
5. *Teacher training* college

---

* The scale points used were: (1) no, (2) maybe, (3) yes.

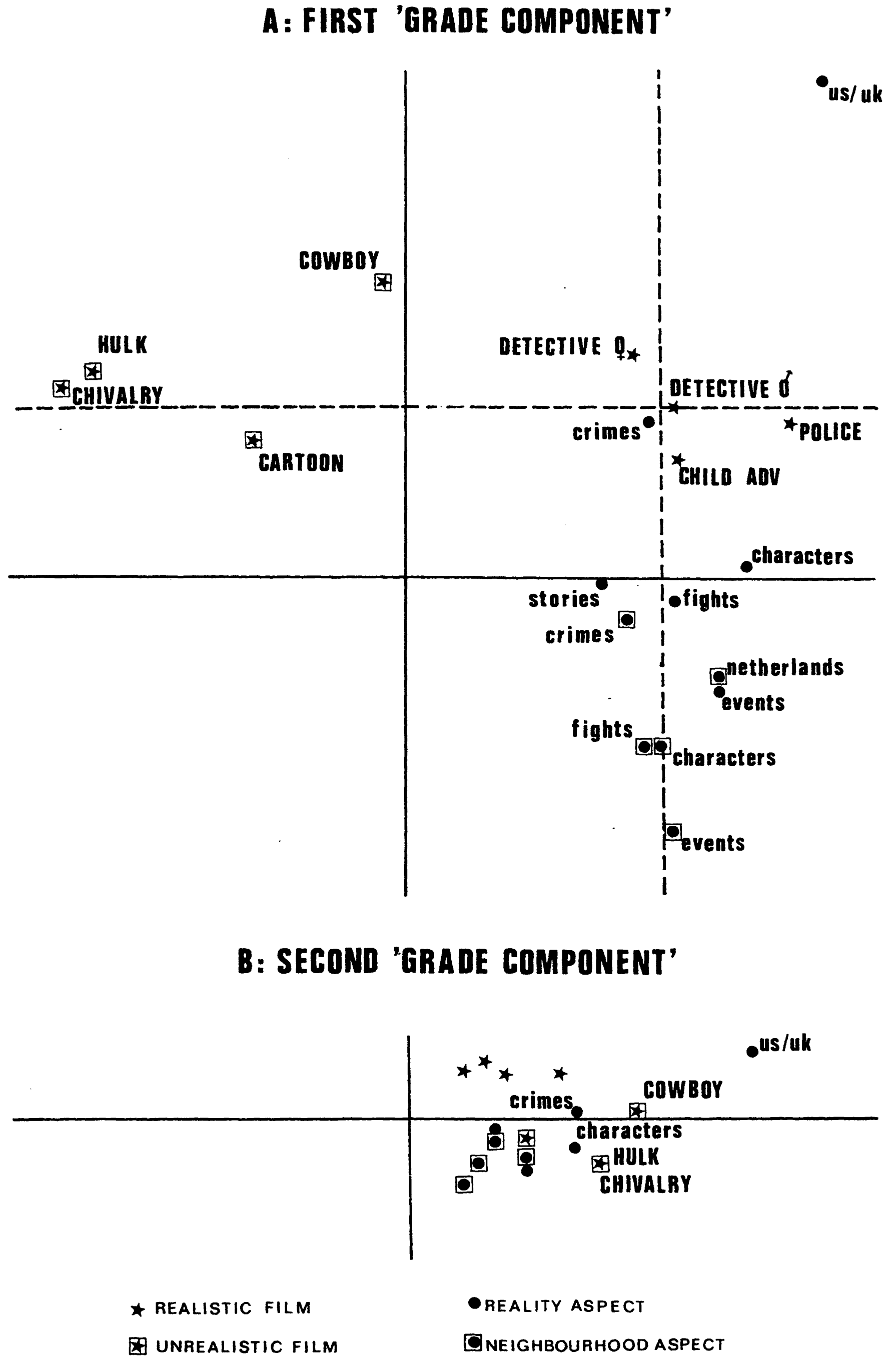
cipal component analysis (see Van der Voort, Vooijs, & Bekker, 1982, p.61-85). Similarly, the film types were also divided into two categories, *realistic* and *unrealistic* films (see again Table 7.2).

Table 7.3 *Perceived reality study: Grade space*

grade	component 1	component 2	scaled component 1	scaled component 2
Third	.41	.51	.37	.11
Fourth	.49	.35	.44	.08
Fifth	.48	.01	.43	.00
Sixth	.47	-.20	.42	-.04
Teacher training	.38	-.76	.34	-.17
length of component	1.00	1.00	$\sqrt{v_1}$	$\sqrt{v_2}$
standardized weight	v_1 = .80	v_2 = .05		

The 'grade space' (Table 7.3) is largely one-dimensional (v_1 = .80) showing the strong similarity in judgement over grades. On the second component (v_2 = .05) the grades were ordered with respect to age. The common structure for all grades is depicted in the *joint plot* (see section 6.10) in Fig. 7.1A. The following conclusions can be drawn from this figure:

1. on all aspects the order of reality of film types is roughly the same (horizontal axis);
2. irrespective of film type the order of the aspects with respect to realism is by and large the same (vertical axis);
3. both divisions show up clearly with a complete separation of the film categories on the horizontal axis, and an overlapping one for aspects on the vertical axis; note that the order (*Events* < *Fights* < *Characters* < *Crimes*) with "<" meaning 'less realistic', is the same for both aspect categories;
4. the scoring on the aspects *events in US/UK* and *crimes in reality* on the one hand, and *events in the neighbourhood* on the other hand, are largely responsible for the neighbourhood-realism distinction.


A: FIRST 'GRADE COMPONENT'

B: SECOND 'GRADE COMPONENT'

Fig. 7.1A and B *Perceived reality study: Joint plots of film types and aspects*

Fig. 7.1B shows the configuration of aspects and film types which corresponds to the ordered grades. From Table 7.3 it can be seen that third graders score more positively and students from the teacher training college more negatively on the second component for grades. Thus positive scores for aspect-film type combinations (i.e. $\sum_{p=1}^{s} g^{*}_{ip} h^{*}_{jp}$, see section 6.10) are associated with positive contributions to the overall score for the third grade and negative contributions for the teacher training grade. In the present example this means that the realism items, especially *events in US/UK*, in combination with unrealistic films, especially *cowboy films*, are responsible for the differentiation between grades. Thus in the third grade especially *events in US/UK, crimes in reality,* and *characters in reality* in unrealistic films are judged to be more realistic than in the middle grades, and the teachers-to-be judged those aspects as less realistic than the middle grades.

7.6 RESIDUAL ANALYSIS FOR PERCEIVED REALITY STUDY

In this section we will give a more detailed discussion of the use of the *analysis scheme* and demonstrate the way in which it can be used.

Distributions of total and residual sums of squares. In Table 7.4 we have presented the distributions of the total and residual sums of squares (or sums of squared residuals), as suggested in the analysis scheme (step 1-A; Table 7.1). For grades we have presented 'regular' tables, while for film types and aspects we have chosen to use Tukey's (1977) stem-and-leaf displays. The advantage of these displays is that they contain the complete information about the data, and show outlying points more clearly.

From Table 7.4 a number of conclusions can be drawn with respect to the data of the study. First the *grades* fit on the whole equally well. It is, however, worthwhile to note that the middle grades (fourth, fifth and sixth) have the lowest SS(Res) and the

Table 7.4 *Perceived reality study:*
Distributions of total and residual sums of squares

Grades	*total sums of squares*		*residual sums of squares*	
	SS(Total$_k$)	SS(Total$_k$)/ SS(Total)	SS(Res$_k$)	SS(Res$_k$)/ SS(Total$_k$)
grade	raw	standardized	raw	Rel. SS(Res$_k$)
G3	79	.18	15	.19
G4	101	.23	14	.14
G5	90	.21	10	.11
G6	91	.21	11	.12
TT	79	.18	16	.21
overall	440	1.00	69	.15
mean	88	.20	13	.15

Film types $)

total sums of squares

	SS(Total$_j$)	
2*	6	Cartoons
3		
4	1568	
5		
6		
7	58	Police ,Hulk
8*	0	Chivalry

Mean = 55

residual sums of squares

	SS(Res$_j$)		Rel. SS(Res$_j$)
5	55	Chivalry, Police	(.07;.07)
6			
7	7	Cartoon	(.27)
8	88	Hulk,Detective ♂	(.11;.19)
9	9	Detective ♀	(.19)
10	0	Cowboy	(.25)
11			
12			
13	3	Child Adventure	(.29)

Mean = 8 Mean = .18

Aspects $)

total sums of squares

	SS(Total$_i$)	
1*	9	Crime. R
2	4889	
3	478	
4	1	Events.N
5		
6		
7		
8		
9		
10		
11*	7	US/UK.R

Mean =44 (Median =31.5)

residual sums of squares

	SS(Res$_i$)	
4	444	
5	5	
6	6	
7	777	
8		
9		
10		
11		
12		
13	3	Real story?

Mean = 6

	Rel.SS(Res$_i$)	
		US/UK.R
.0*	49	Events.R
.1	3557	
.2	234	
.3		
.4		
.5*	5	Real story?

Mean =.19

$) stem-and-leaf displays (Tukey, 1977: Chapter 1); '*' indicates the leaf is the next digit, i.e. 1* 9 = 19; in unstarred displays leaf = stem extreme grades (third grade and teaching training college), the largest ones, showing that the joint plots of aspects and films represent the views of the middle grades better than those of the extreme grades.

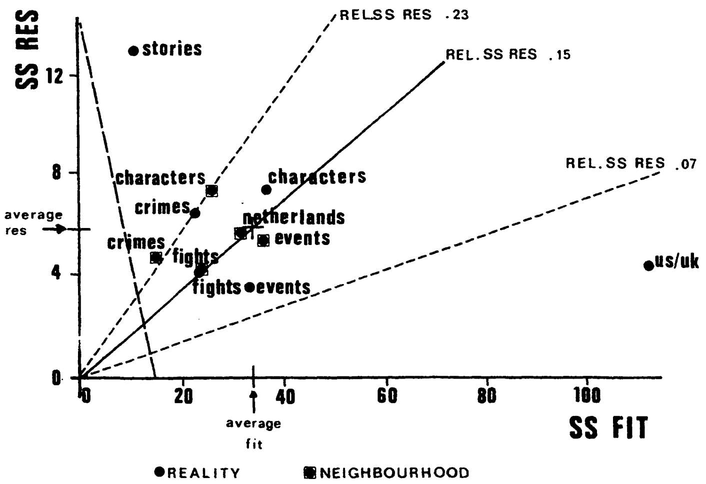
With respect to *films*, it can be noted that the total sums of squares for *cartoons* is small, in other words, the variation around the overall mean value is small. Apparently this type of film is judged alike over all aspects by all grades with an about average reality value. On the other hand *chivalry films* are judged rather differently either over aspects or grades or both. The SS(Res) of *children's adventure films* is larger than the other film types, and thus this film type does not fit as well.

Finally, with respect to *aspects*, *events in US/UK* has a large SS(Total), 26% of the overall total sum of squares being associated with this aspect alone. According to our analysis scheme, one might consider eliminating this aspect and redoing the analysis to assess its influence on the overall configuration. This was done and the structures of Fig. 7.1 still showed up, though not as clearly. *Based on a real story* does not fit well in the overall pattern with about 55% of its total variation not accounted for by the model, compared to a maximum of 24% for any of the other aspects.

Analysis-of-variance decomposition of sums of squares. In the previous paragraph we have looked at the three modes separately, and not comparatively. It is, however, useful to investigate the relative sizes of the differences between the sums of squares of the modes. One way of doing so is to make an analysis-of-variance decomposition of the squared residuals. The relevant quantities for the present example are given in Table 7.5. As the constant is the estimate of the amount the squared residuals (ε^2_{ijk}) differ from zero, its contribution to the total variation of the squared residuals is an indication for the overall size of these residuals. Over and above this contribution of 27%, the sizes for the main effects are only marginally important: 5% for *aspects*, 2% for *film types*, and 1% for *grades*. The only sizeable contribution which might be worth investigating further seems to be that of *aspects*. From our previous analysis we know that *based on a real story* is largely responsible for the variation in the residual sums of

Table 7.5 *Perceived reality study: analysis-of-variance decomposition for squared residuals*

source	*formula*	*interpretation*	*SS*	*%*
Constant	$\sum_{i=1}^{\ell}\sum_{j=1}^{m}\sum_{k=1}^{n}\{\frac{1}{\ell mn}\sum_{i'=1}^{\ell}\sum_{j'=1}^{m}\sum_{k'=1}^{n}\varepsilon^2_{i'j'k'}\}^2 =$ $\ell mn\{\frac{1}{\ell mn}\ SS(Res)\}^2 = \frac{1}{\ell mn}\{SS(Res)\}^2$	variation estimate for deviations from zero of the squared residuals, or variation due to the mean squared residual	10.1	27
Aspects	$\sum_{i=1}^{\ell}\sum_{j=1}^{m}\sum_{k=1}^{n}\{\frac{1}{mn}\sum_{j'=1}^{m}\sum_{k'=1}^{n}\varepsilon^2_{ij'k'} - \frac{SS(Res)}{\ell mn}\}^2 =$ $\sum_{i=1}^{\ell}\sum_{j=1}^{m}\sum_{k=1}^{n}\{\frac{SS(Res_i)}{mn} - \frac{SS(Res)}{\ell mn}\}^2 =$ $\frac{1}{mn}\sum_{i=1}^{\ell}\{SS(Res_i) - \frac{1}{\ell}SS(Res)\}^2$	variation estimate for deviations of the residual sum of squares of the aspects from the mean squared residual	1.8	5
Film types	$\sum_{i=1}^{\ell}\sum_{j=1}^{m}\sum_{k=1}^{n}\{\frac{1}{n\ell}\sum_{k'=1}^{n}\sum_{i'=1}^{\ell}\varepsilon^2_{i'jk'} - \frac{SS(Res)}{\ell mn}\}^2 =$ $\frac{1}{n\ell}\sum_{j=1}^{m}\{SS(Res_j) - \frac{1}{m}SS(Res)\}^2$	idem for film types	0.9	2
Grades	$\sum_{i=1}^{\ell}\sum_{j=1}^{m}\sum_{k=1}^{n}\{\frac{1}{\ell m}\sum_{i'=1}^{\ell}\sum_{j'=1}^{m}\varepsilon_{i'j'k} - \frac{SS(Res)}{\ell mn}\}^2 =$ $\frac{1}{\ell m}\sum_{k=1}^{n}\{SS(Res_k) - \frac{1}{n}SS(Res)\}^2$	idem for grades	0.3	1
Remainder (ANOVA-Res.)	$\sum_{i=1}^{\ell}\sum_{j=1}^{m}\sum_{k=1}^{n}\{\varepsilon^2_{ijk}\}^2$ - {above contributions}	remaining variation after fitting the three-way main effects model	24.0	65
Total	$\sum_{i=1}^{\ell}\sum_{j=1}^{m}\sum_{k=1}^{n}\{\varepsilon^2_{ijk}\}^2$		37.1	100


ASPECTS

———— line connecting (0,0) and (av. SS(Fit),av.SS(Res)). It separates the well-fitting points (below the line) from the ill-fitting points (above the line).

------ line of equal relative SS(Res).

— — — line at -45° with the x-axis:locus of points with equal SS(Tot).

·········· locus of points with decreasing relative SS(Fit).

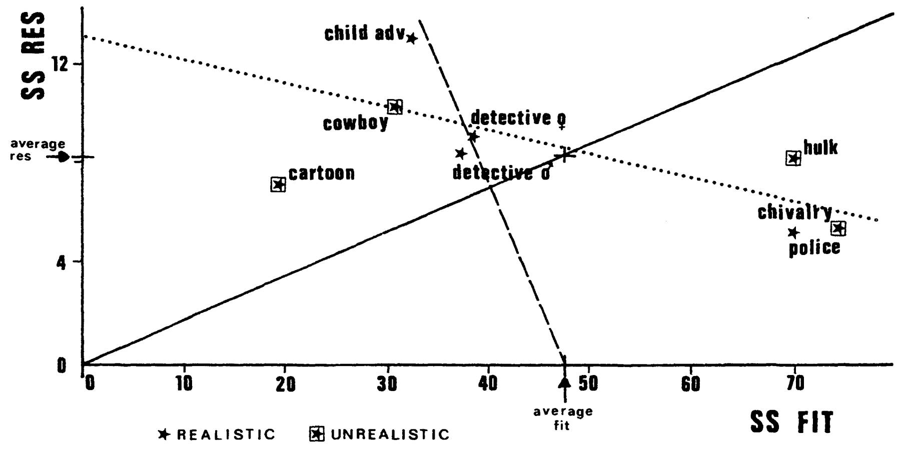

FILM TYPES

Fig. 7.2 *Perceived reality study: Sums-of-squares plots of aspects and film types* squares for aspects, and that to a lesser extent the same can be said for *children's adventure stories* with respect to films. One could attempt at this stage of analysis to look at interactions after all the ANOVA-residuals account for 65% of the sum of squares of the ε^2_{ijk}, but we prefer to look at the individual residuals via *Step 2* of Table 7.1.

Sums-of-squares plots. To assess the quality of the fit of the elements of a mode, it is particularly useful to look at the residual sums of squares in conjunction with the total sums of squares of the elements. One way to do this is to investigate the *relative residual sums of squares* [Rel.SS(Res) = SS(Res)/SS(Total)]. Or one could look at the *Residual/fit ratio*, i.e. SS(Res)/ SS(Fit), from which the relative performance of an element can be gauged. Of course, the two measures carry the same information.

The SS(Res) and the SS(Fit) as well as their relationships can be shown directly in a so-called *sums-of-squares plot*. For the aspects of the *Perceived reality study* such a plot is shown in Fig. 7.2A. By plotting the sums of squares directly, rather than the relative sums of squares, the total sums of squares are also contained in the plot and unusually large elements can be spotted immediately. Moreover, it can be seen if the larger SS(Fit) resulted only from larger total sums of squares as is to be expected from least squares procedures. This will be evident in the plot when elements lie on a line with an angle acuter than -45° (see Fig. 7.2B). Furthermore, when the variations (variances or total sums of squares) of the elements have been equalized, this is evident from the arrangement of the elements on a line at an angle of -45° with the positive x-axis (see Fig. 7.3 for a sums-of-squares plot of the standardized variables in the *Hospital study* (Chapter 13). Note that because the axes represent sums of squares, the total sums of squares are obtained by directly adding the x-value and the y-value.

Another interesting feature of these plots is that they show which elements have equal residual sums of squares with different total sums of squares. In other words, it becomes possible to separate points which have large residual sums of squares because they

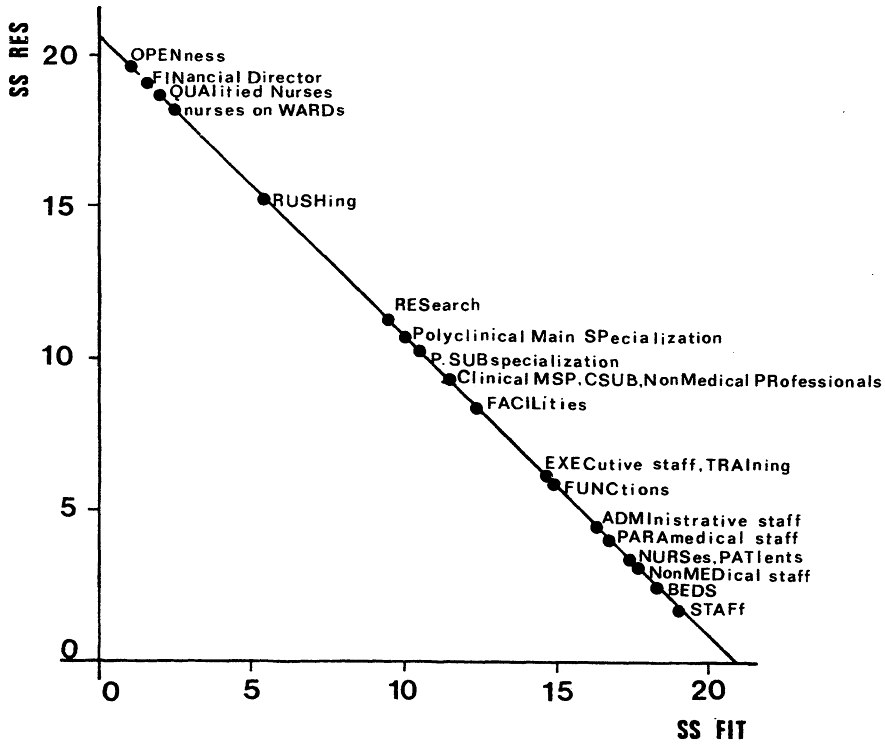


Fig. 7.3 *Hospital study: Sums-of-squares plot - variables*

do not fit well, from the points which have a large SS(Res) because they have a large total sum of squares. Without a residual analysis it is uncertain whether a point in the middle of a configuration on the first principal components is an ill-fitting point or just a point with little overall variation.

Finally, by drawing the line through (0,0) and (av.SS(Fit), av. SS(Res)), and appropriate similar lines above and below it, something akin to confidence bands can be constructed around the former line to assess the extremity of certain elements. The lines are the loci of points with equal Res/fit ratios (or equal Rel. SS(Res), which comes down to the same thing). A guideline for what is 'appropriate' in this case, i.e. how much the individual element may deviate in relative residual sum of squares from the overall one, has not been developed yet. The lines in Fig. 7.2A are the lines "overall Rel.SS(Res) ± .08" and they seem reasonable bounds for the present data.

With respect to the aspects, both the Rel.SS(Res) of *events in US/UK* and *based on a real story* deviate considerably from the overall Rel.SS(Res) or Res/fit ratio. We already noticed this above, but the advantage of the plot is that these points can be also easily recognized with a large number of elements.

Individual residuals (unstructured approach). In the *Perceived reality study* the largest anomaly is the very large total sum of squares of *events in US/UK*. The not entirely acceptable Rel. SS(Res)'s for *children's adventure films* and *based on a real story* suggest that it might be worthwhile to look at the individual residuals to gain some further insight.

According to our *analysis scheme* of Table 7.1, we should first have a look at the distribution of the residuals. The total sample consists of 440 points (11×8×85), and if Gnanadesikan's remark (1977, p.265) is true that residuals tend to be 'supernormal', or at least more normal than the data, it seems worthwhile to compare both the data and the residuals with the normal distribution (Fig. 7.4). Most standard statistical packages, such as SPSS and BMDP, have options to produce normal probability plots, generally accompanied by detrended normal plots (i.e. plots of deviations from the expected normal distributions against the residuals). Bock (1975, p.155-160) and Meerling (1980, p.131-137) provide discussions of these plots and their interpretation. Comparison with Bock's figures show the data to be reasonably normally distributed (except, maybe, for a few points at the far ends of the distribution) with a slight right skew. The residuals bear some resemblance to a leptokurtic distribution with a higher concentration around the mean (here: about zero) and heavier tails than the normal density. The higher concentration around the mean of a leptokurtic density is rather attractive for residuals, as this is just what is desired; the heavier tails, however, are somewhat less desirable as they indicate the presence of outliers. In the present data there is some evidence that the residuals are more normal than their data.

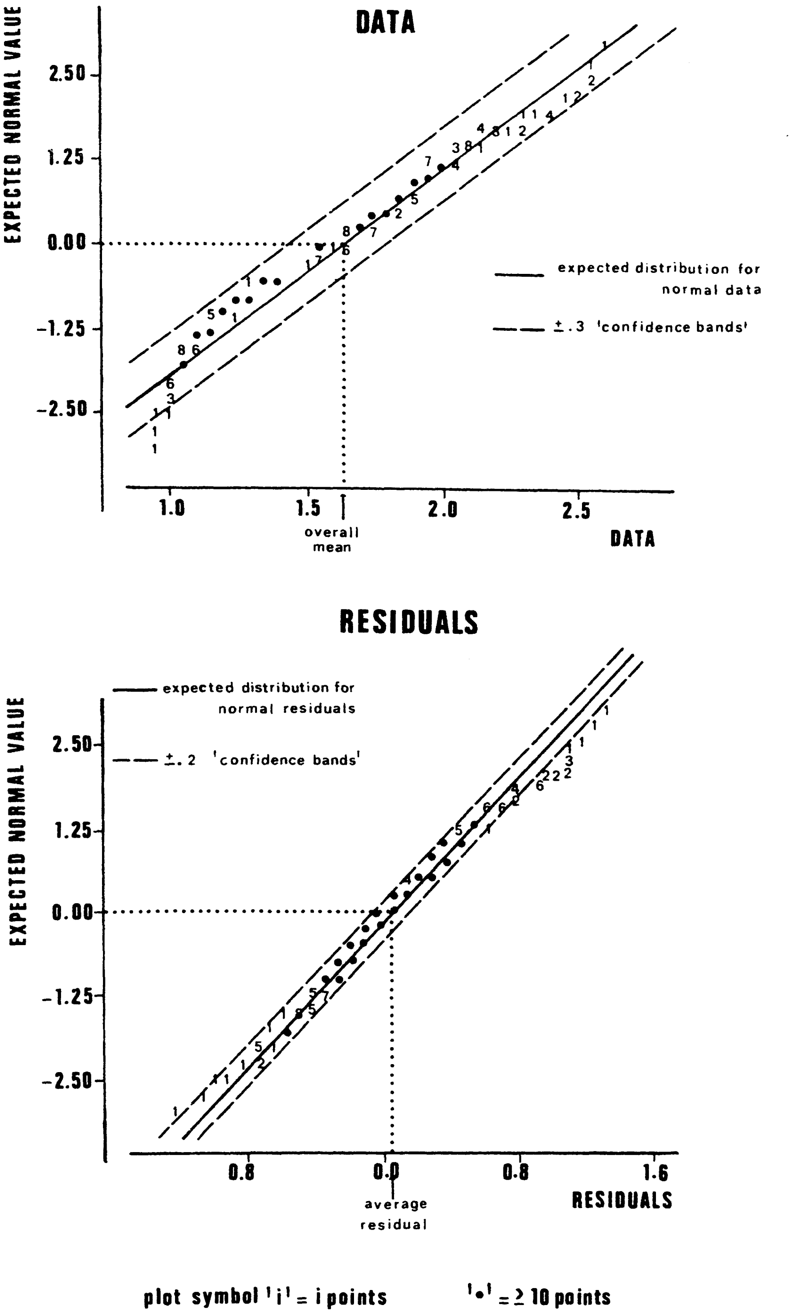


Fig. 7.4 *Perceived reality study: Normal probability plots*

Table 7.6 *Preceived reality study: extreme residuals*

Grade£)	aspect$)	film type	residual*
		right-hand tail	
4	characters -N	children's adventures	1.30 o
5	characters -N	children's adventures	1.17 o
4	fights -N	children's adventures	1.09 o
4	real story? -R	chivalry	1.05 +
3	characters -R	cowboy	1.03
3	real story? -R	chivalry	1.02 +
4	characters -R	cowboy	1.02
5	events -N	female detectives	1.02
T	crimes -R	male detectives	1.01
3	in US/UK -R	children's adventures	.97 o
6	events -R	hulk-like	.96
T	in Netherlands -N	male detectives	.90
3	events -N	female detectives	.87
6	fights -R	children's adventures	.82 o
T	real story? -R	chivalry	.82 +
4	events -R	cowboy	.81
5	characters -N	cartoons	.81
4	fights -R	cowboy	.80
		left-hand tail	
T	events -R	children's adventures	- .81 o
T	in US/UK -R	children's adventures	- .82 o
3	crimes -N	cowboy	- .83
3	in Netherlands -N	cowboy	- .84
6	real story? -R	hulk-like	- .96 +
3	real story? -R	cartoons	-1.03 +
5	real story? -R	hulk-like	-1.11 +
T	characters -R	female detectives	-1.25

£ T = students from teacher training college
$ N = neighbourhood, R = reality
* + = residual involving *based on a real story*
o = residual involving *children's adventure film.*

Both at the lower and upper end of the density of the residuals there are some irregularities that might be worth investigating; they are listed in Table 7.6. There does not seem to be very much system in the larger residuals, except that *children's adventure films* (indicated in Table 7.6 with a "o") produces some big residuals, which is consistent with its position in Fig. 7.2B. The first two positive residuals indicate, for instance, that grades 4 and 5 considered the occurence of characters in this film type far more likely in the neighbourhood than the model would predict. The large residual sum of squares for *based on a real story* is the result of various large, both positive and negative, residuals in connection with unrealistic films (indicated in Table 7.6 with a "+"). On the whole the larger residuals are a rather mixed bag, in which various individual residuals are open to interesting, but maybe spurious, interpretations. A nice case in point is the relative disbelief of the students from the teacher training college in the reality of the female detectives (probably especially in "Charlie's Angels"). As suggested in our *analysis scheme* the residual-versus-data plots were also inspected, but no systematic trends were found.

In summary it can be said that the analysis of the residual sums of squares and the residuals themselves provided some useful insight in the quality of the solution found for the data in the *Perceived reality study*. It pointed towards the large influence of the perceived reality of *events in the US/UK* (for Dutch children!). Furthermore, both *children's adventure films* and *adventures in reality* (especially for realistic films) did not fit the model as well as the other films and aspects respectively. However, it is clear that the data set as a whole, and the data points individually, conformed quite well to the model fitted.

7.7 THREE-MODE ANALYSIS AND THREE-WAY ANOVA DECOMPOSITION

The structured approach towards analysing residuals in this chapter is in some sense the counterpart of that in the previous one (especially section 6.4). There we discussed fitting a three-way analysis-of-variance model to the data, and investigating the residuals with three-mode principal component analysis. In this chapter we proposed, among other things, to look at residuals from three-mode component analysis by an analysis of variance decomposition of the squared residuals, i.e. by looking at the residual sums of squares for each of the three modes in the same way as looking at main effects.

The duality extends to the kind of data and their appropriate analysis. With multiresponse data without independent variables, it is more appropriate to perform principal component analysis first. The residuals from this analysis can then be treated as a set of measures on one dependent measure, i.e. *badness-of-fit*, in an $\ell \times m \times n$ completely crossed factorial design without replications, and thus balanced as well. On the other hand a set of original observations on one dependent variable from an $\ell \times m \times n$ factorial design is best handled within an analysis of variance framework. Augmenting such an analysis by a principal component analysis of the residuals may shed light on correlations of residuals of the combined original variables, or indicate inadequacies in the fit of the dependent variable by the (latent) predictors or design variables (see also Gnanadesikan, 1977, p.259).

A nice aspect of the data from the *Perceived reality study* is that they may be used in both ways, which is the reason why we classified them as anova-data in section 6.2. If one considers the data as scores on one single dependent variable - *perceived reality* - the ANOVA-first approach is appropriate. If one looks upon the data as multiresponse data, the PCA-first approach (followed in this chapter) is the logical choice. Of course, the purpose of the analysis in the first place determines what is appropriate.

As a final remark about the differences of the two approaches, it should be mentioned that from a statistical point of view the analysis of variance as a first step seems to be a more manageable situation. For instance, as Mandel (1969) showed for the two-way case, tests can be derived for the principal components after the analysis of variance, of course under certain assumptions. Using an analysis of variance after principal component analysis runs into difficulties in this respect, as the observational units (usually subjects) generally have residual scores in an entire matrix, rather than in just one cell, thus suggesting a factorial design with repeated measures on two factors without replications, which might be difficult to handle. But even if this were manageable (as in principle it is), the (possible) input scaling and the principal component analysis itself introduce additional dependencies which might be extremely hard to define and handle.

III

APPLICATIONS

SUMMARY

After the general and theoretical discussion in Part II of issues connected with applying three-mode principal component analysis, a number of applications are presented in *Part III*. The applications were chosen to be representative for a particular class of applications. In this way they may serve as a guide for the analysis of similar data.

Chapter 8 (Attachment study) contains a relatively detailed account of a data set from development psychology. It deals with the reaction of young children to a standardized procedure designed to measure the attachment of those children to their mother figure. In a way this set is typical of the kind of data for which three-mode principal component analysis can be useful.

Chapter 9 (Triple personality study) contains an example from research using the semantic differential technique with a special focus on individual differences. In addition, the concept-scale relationships are a major focus of the enquiry. The data used have been produced by a single person in three different personalities, Eve White, Eve Black, and Jane. The handling of the data may stand as an example for many studies which include comparable rating scales or tests.

Chapter 10 (ITP study) contains an example with asymmetric similarity data on implicit theories of personality, but the analysis of the data is not presented in full. The focus is on an interpretation aid, called the theoretical subject, who turns out to be helpful in assessing subject spaces.

Chapter 11 (Cola study) contains a re-analysis of data from sensory perception, and investigates whether direct comparisons of similarities between colas have anything in common with ratings of the same colas on a number of adjectives. The data set, at least the similarity part of it, is a typical example of the kind of data usually treated by techniques for individual differences scaling.

An extensive comparison with such techniques is given as well.

Chapter 12 (Four ability-factor study) is an example of the use of three-mode principal component analysis for the analysis of correlation matrices. Typically cross-sectional data can be handled in this way, as well as a re-analysis of published material not available as raw data. The example is taken from the field of intelligence testing, and the structure of a series of tests is investigated for various age groups of normal and retarded children.

Chapter 13 (Hospital study) contains an example from organizational sociology. In this study the structura: organization of Dutch hospitals is investigated over a number of years. Various issues with respect to the analysis of multivariate longitudinal data are treated, in particular the relationship between three-mode analysis and autoregressive models.

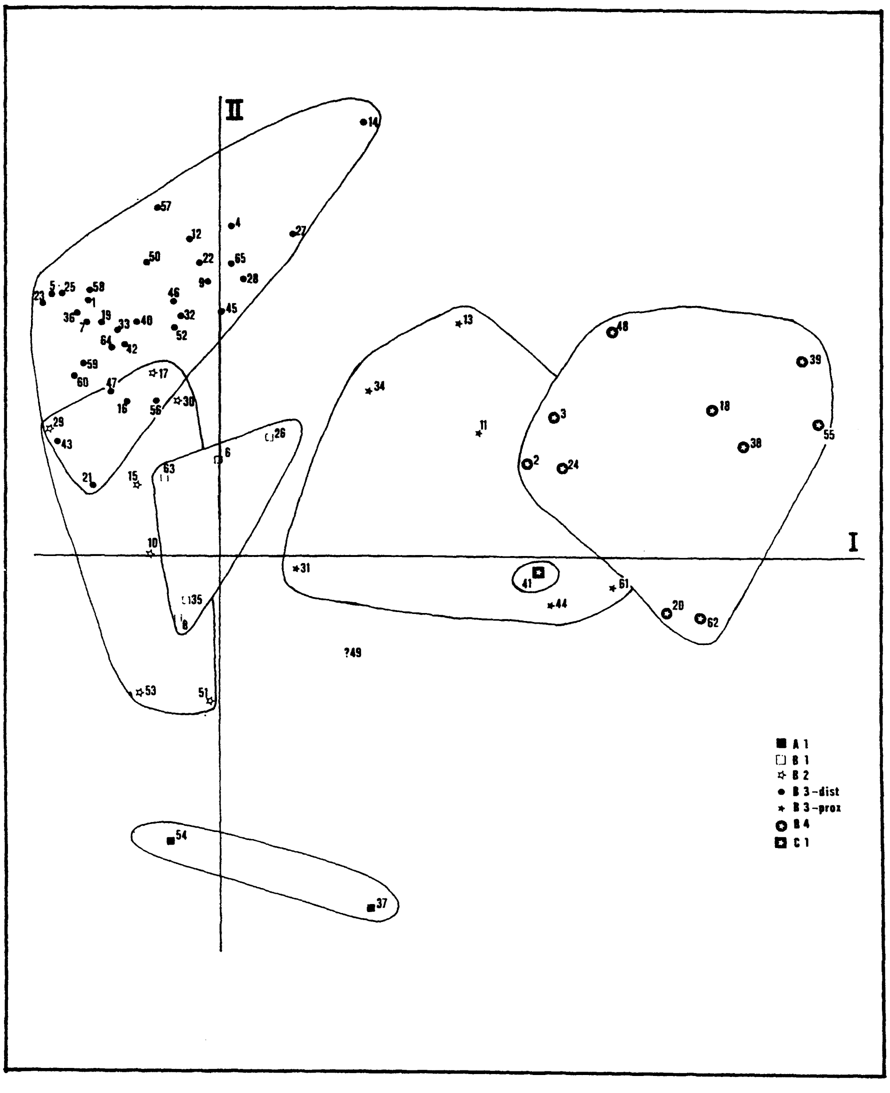
Chapter 14 (Learning-to-read study) contains an extension of Tucker's work on generalized learning curves by analyzing the data of a study investigating the process of learning to read. A brief comparison is made with results from linear logistic modelling on the same data.

Chapter 15 (Leiden electorate study) contains an example rather different from the rest in that it deals with counted rather than measured data. The data consist of the results from three different elections held in Leiden, The Netherlands. The three-way contingency table is analyzed by loglinear methods to determine the relevant interactions, and subsequently the residuals are analyzed with three-mode principal component analysis. The procedure used is a three-mode extension of correspondence analysis.

STANDARD
THREE-MODE
DATA

8

attachment study



8.1 DESIGN AND DATA DESCRIPTION

In this chapter we present a relatively straightforward analysis of data collected by Goossens (Note 1) on the reactions of two-year old children to a stranger and to their mothers in an unfamiliar environment within the context of a standardized observation procedure called the *Strange Situation (Patterns of Attachment (POA),* Ainsworth, Blehar, Waters, & Wall, 1978). The practical aspects and theoretical considerations which form the foundation of the strange situation are covered in many publications including the above, as the measurement procedure has become a standard one in developmental psychology. Therefore we will not dwell in detail on the strange situation, but only treat those aspects necessary to an understanding of the data and their analysis.

In the course of the strange situation the child is subjected to increasingly stressful circumstances (i.e. arrival of a stranger, leaving of the mother, being left alone) in order to elicit 'attachment behaviours'. *Attachment* itself is defined as "the affectional bond or tie that an infant forms between himself and his mother figure - a bond that tends to be enduring and independent of specific situations", and *attachment behaviours* are defined as "the class of behaviours that share the usual or predictable outcome of maintaining a desired degree of proximity to the mother figure" (Ainsworth et al., 1978, p. 302).

As Ainsworth et al. point out (p.33), the sequence of episodes is very powerful both in eliciting the expected behaviours, and in highlighting individual differences. The major purpose of the procedure is to assess the quality of the attachment relationship of a child to its mother-figure. A summary of the procedure is given in Table 8.1A. The major types of attachment are *secure attachment* (B-children), *anxiously resistant attachment* (C-children), and *anxiously avoidant* attachment (A-children).

Table 8.1 *Attachment study: Description of strange situation, interactive scales, and classification system*

A. *Strange situation,* (*POA,* p.37)

Episode	Persons	Duration	Brief description of action
1	mother, child, observer	30 secs.	Observer introduces mother and baby to experimental room, then then leaves.
2	mother, child	3 min.	Mother is non-participant while child explores; if necessary, play is stimulated after two minutes
3	stranger, mother, child	3 min.	Stranger enters. First minute: stranger silent. Second minute: stranger converses with mother. Third minute: stranger approaches child. After three minutes mother leaves unobtrusively.
4	stranger, child (S4)	3 min. or less 1)	First separation episode. Stranger's behaviour is geared to that of the child.
5	mother, child (M5)	3 min. or more 2)	First reunion episode. Mother greets and/or comforts child, then tries to settle it again in play. Stranger leaves unobtrusively in the meantime. Mother leaves saying "bye bye".
6	child alone	3 min. or less 1)	Second separation episode.
7	stranger, child (S7)	3 min. or less 1)	Continuation of second separation. Stranger enters and gears behavior to that of the child.
8	mother, child (M8)	3 min.	Second reunion episode. Mother enters, greets child, then picks it up. Meanwhile stranger leaves unobtrusively.

1) Episode is curtailed if the child is unduly distressed.
2) Episode is prolonged if more time is required for the child to become reinvolved in play.

Table 8.1 *(cont'd)*

B. *Interactive scales* (*POA*, p.53, 54)

- Proximity (or contact)-seeking (PROX): a measure for the degree of active initiative a child shows in seeking physical contact with or proximity to an adult.
- Contact maintaining (CM) : a measure for the degree of active initiative a child exerts in order to maintain physical contact with a person, once such contact is achieved.
- Resistance (RES) : a measure for the degree of angry and/or resistant behaviour to an adult. It is shown by physically rejecting an adult who tries to come into contact or initiate interaction with the child.
- Avoidance (AVOI) : a measure for the degree of avoiding proximity and interaction with an adult, for instance by ignoring or looking away.
- Distance interaction (DI) : a measure for the degree in which a child interacts with an adult from a distance, for instance, by showing toys and talking.

C. *Ainsworth Classification Categories* (based on *POA*, p. 59-63; Sroufe & Waters, 1977)

Behaviour towards the mother

	PROX	CM	RES	AVOI	DI	most salient feature	behaviour towards stranger
A1	-	-	-	++	-	disinterested	treatment more or less like mother
A2	+(+)	-	(+)	++	-	mixed feelings	
B1	(+)	-	-	-	++	secure	friendly towards stranger but mother is clearly preferred and sought after
B2	+(+)	(+)	-	(+)	+(+)	secure	
B3	++	++	-	-	-/++	very secure	
B4	++	++	(+)	-	-	secure	
C1	++	++	++	-	-	angry ambivalent	treatment more or less like mother
C2	(+)	(+)	++	-	(+)	passive	

- low; (+) low to moderate;
+ moderate; +(+) moderate to high;
++ high.

POA: Patterns of attachment, Ainsworth et al. (1978)

Ainsworth et al. (1978, Ch.3) have developed a more detailed classification system, which is presented in Table 8.1C. The classifications of the children are made by trained judges on the basis of the children's scores on so-called *interactive scales* which range from 1 to 7. The child's behaviour corresponding to each of the seven categories has been explicitly defined, and can be summarized as going from 1 (virtually non-existent) to 7 (very often, very intense). The scores are awarded by trained observers, while viewing videotapes of the strange situation. In the present analysis the following scales were used: *proximity seeking* (PROX), *contact maintaining* (CM), *resistance* (RES), *avoidance* (AVOI), and *distance interaction* (DI) (see Table 8.1B).

The data consisted of observations on 65 two-year old children on the 5 interactive scales during 4 episodes (S4, M5, S7, M8), where S indicates the presence of the stranger and M that of the mother. Details on the data and the reasons for discarding the earlier episodes can be found in Goossens (Note 1). One might argue that a three-mode analysis is not a proper technique for these data, as for instance proximity seeking towards the stranger might not be the same variable as proximity seeking towards the mother. Moreover, the relationships between the scales in the stranger episodes might be different from those in the mother episodes. However, as the basic purpose of the strange situation is to assess children on the basis of their reactions to the entire strange situation, and not to specific parts of it, it seems justified to treat a scale as the same variable regardless of the adult towards whom the behaviour is directed.

Before analysis, the overall scale means were removed, i.e. the scales were centred over all children-episode combinations (j-centring - see section 6.5). No equalization of variances was performed. This decision was based on the consideration that not the overall scoring levels of the children on the interactive scales were of interest, but the individual differences between children. This centring ensures that the meaningful differences in scoring levels between episodes which carry important information are retained. A disadvantage of using the mean values for generalization is that they are sample dependent. For more extensive studies some standard norm for centring the scales should be devised.

8.2 ANALYSIS AND FIT

The main analysis reported here is a Tucker3 (T3) analysis with two components each for the first mode (episodes), second mode (interactive scales), and third mode (children). It will be referred to as the *2x2x2-solution*, and will be compared with a *3x3x3-solution* on the same data. At times we will also refer to a Tucker2 (T2) analysis with two components for the first two modes, or the *2x2-solution*.

Table 8.2 shows that with an increasing number of components the fit increases, but that the increase in fit in going from the 2x2x2-solution (fit = .59) to the 3x3x3-solution (fit = .68) involves estimating an additional 93 parameters. At least three-fifth of the variation in the (j-centred) data is accounted for by the three-mode model. Considering the relative difficulty of reliably measuring children's behaviour, and the variability inherent in it, this seems quite satisfatory.

Table 8.2 *Attachment study: Characteristics of the solutions*

	T3 *2x2x2*	T3 *3x3x3*	T2 *2x2*
Standardized total sum of squares - SS(Total)	1.00	1.00	1.00
Approximation of SS(Fit) from separate PCA			
on mode 1	.77	.91	.77
on mode 2	.83	.92	.83
on mode 3	.63	.71	-
Fitted sum of squares from simultaneous estimation - SS(Fit)	.59	.68	.67
Residual sum of squares from simultaneous estimation - SS(Res)	.41	.32	.33
Improvement in fit compared to initial configuration	.03	.01	.001
Parameters to be estimated	156	249	278

When using the Tucker2 model, i.e. computing only components for episodes and interactive scales, a better overall fit is possible than with the Tucker3 model with the same number of components

(.67 for the 2x2-solution versus .59 for the 2x2x2-solution). But due to leaving the third mode unreduced there are more parameters in the former model (278 versus 156). Comparing the two T3-solutions, it is difficult to decide which is the 'best' solution to look at in detail. No goodness-of-fit tests are available, and, in addition, it seems largely a content-specific problem in how much detail one wants to describe the relations.

8.3 CONFIGURATIONS OF THE THREE MODES

The (common) component spaces for each mode are given in Table 8.3A,B,C. In Fig. 8.1 the components for scales and episodes are plotted, and in Fig. 8.2 those for the children. In Fig. 8.1 A,B, but not in Fig. 8.2, the components have been multiplied by the square root of their component weights, so that the plots reflect the relative importance of the axes (see section 6.8).

The general remark can be made that on the whole the choice of a particular solution is not very crucial with respect to interactive scales and episodes. The first two components of both the scale space and the episode space are the same within reasonable bounds (roughly ± .05; the order is preserved in all but two cases).

Table 8.3 *Attachment study: Component spaces*

A. *Episodes* (mode 1)

nr.	adult		T3: 2x2x2		T3: 3x3x3			T2: 2x2	
			E1	E2	E1	E2	E3	E1	E2
4	stranger	S4	.26	-.44	.25	-.37	.45	.26	-.45
5	mother	M5	.47	.25	.52	.28	.68	.48	.27
7	stranger	S7	.38	-.77	.41	-.80	-.23	.44	-.73
8	mother	M8	.75	.39	.71	.38	-.53	.71	.43
component weight (λ_p)			.37	.22	.41	.21	.07	.42	.25

Labels for components:	
	E1, stress of situation
	E2, mother versus stranger
	E3, early versus late

Table 8.3 *(cont.d)*

B. *Interactive scales* (mode 2)

Scales		T3: 2x2x2		T3: 3x3x3			T2: 2x2	
		S1	S2	S1	S2	S3	S1	S2
Proximity seeking	PROX	.32	.69	.37	.68	.04	.35	.67
Contact maintaining	CM	.26	.35	.26	.34	.14	.28	.34
Resistance	RES	.33	-.41	.30	-.39	.85	.30	-.39
Avoidance	AVOI	.27	-.48	.25	-.50	-.46	.25	-.53
Distance interaction	DI	-.81	.07	-.80	.12	.24	-.80	.10
component weight (μ_q)		.37	.22	.43	.24	.02	.40	.27

Labels for components:
S1, intensity of reaction
S2, security seeking
S3, interest in adult

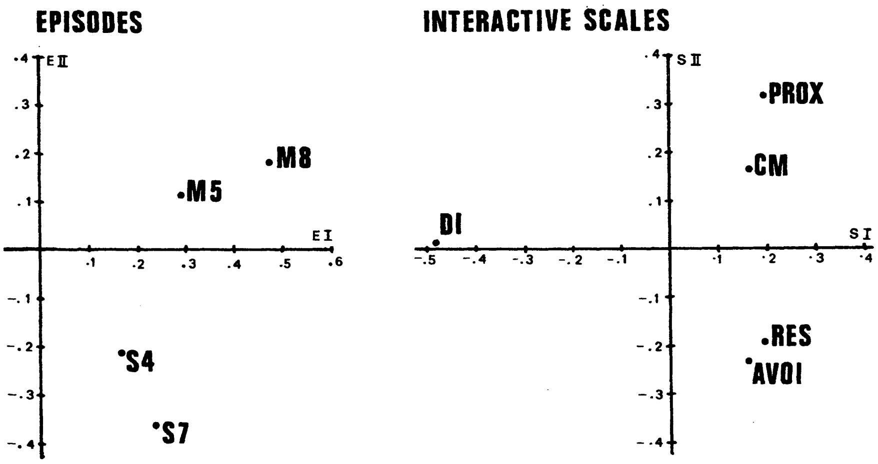


Fig. 8.1 *Attachment study: Component spaces (scaled)*

A point which should be made at the outset of the interpretation is that it is rather difficult to link the details of our results to those in *POA* as the latter refer mainly to one-year olds, and Goossens's study deals with two-year olds. Previous research (summarized in *POA*) shows that the reaction of older children in the strange situation is different from that of the one-year olds it has been validated for (see also Goossens, Swaan, Tavecchio, Vergeer, & Van IJzendoorn, 1982).

One of the aims of the present analysis is to investigate how individual differences between children can be traced back to their different behaviour in the various episodes, on the basis of the interactive scales. These results will then be compared with the classification (sub)categories resulting from the scoring instructions in *POA* (see Goossens, Note 1). One qualification should be made in advance, as the research project from which these data have been derived is not yet finished. Both the scoring of the data and the results presented here should be seen as a first exploration, not yet as definite. The final version will be published elsewhere at a later date.

Episodes. With only four episodes there is really no need to label the axes, but for further reference we will try to name them anyway (see section 6.8). The first axis (E1) reflects the overall variability of the scores in the episodes, and it does not seem unreasonable to associate increasing variability with greater *stress* put onto the child. The second axis (E2) contrasts the behaviour towards the *mother* and that towards a *stranger*. The third axis (E3), finally, contrasts the *early* and *late* episodes, i.e. those episodes before and after episode 6, in which the child has been left alone. If desired two oblique axes could be chosen as well, one for the mother episodes, one for the stranger episodes.

Interactive scales. The first axis (S1) reflects the overall variability of the children-episode combinations around the overall scale mean. This variability is approximately equal for PROX, CM, RES and AVOI, and considerably larger for DI. High scores on distance interaction reflect an opposite reaction compared to high scores on the other scales, and the same holds for low scores. This is to be expected as proximity seeking more or less precludes distance interaction and vice versa. The special position of distance interaction has been noted before, and a number of research-

Table 8.3 *(cont'd)*

C. *Children* (mode 3)

nr.	ACC	C1	C2
55	B4	.34	.08
39	B4	.33	.12
38	B4	.30	.07
18	B4	.28	.09
62	B4	.27	-.03
20	B4	.25	-.03
48	B4	.22	.14
61	B4	.22	-.02
24	B4	.20	.05
3	B4	.19	.08
44	B3	.19	-.02
2	B4	.18	.06
41	C1	.18	-.01
11	B3/4	.15	.07
13	B3	.14	.14
34	B3	.08	.10
14	B3	.08	.26
57	B3	-.04	.21
4	B3	.01	.20
12	B3	-.01	.19
27	B3	.04	.19
22	B3	-.01	.18
50	B3	-.04	.18
65	B3	.01	.18
28	B3	.02	.17
9	B3	-.00	.17
25	B3	-.09	.16
5	B3	-.09	.16
58	B3	-.07	.16
46	B3	-.03	.15
1	B3	-.07	.15
36	B3	-.08	.14
23	B3	-.10	.15
45	B3	-.00	.15
52	B3	-.03	.14
32	B3	-.02	.14
40	B3	-.05	.14
19	B3	-.07	.14
7	B3	-.07	.14
33	B3	-.06	.13
64	B3	-.06	.13
42	B3	-.05	.13
59	B3	-.08	.12
60	B3	-.08	.11
17	B2	-.04	.11
47	B3	-.06	.10
30	B2	-.02	.09
56	B3	-.04	.09
16	B3	-.05	.09
29	B2	-.09	.08
43	B3	-.09	.07
26	B1	-.03	.07
6	B1	-.00	.06
63	B2	-.03	.05
15	B2	-.05	.04
21	B3	-.07	.04
10	B2	-.04	.00
31	B3	.04	-.01
35	B1	-.02	-.02
8	B1	-.02	-.04
49	?	.07	-.06
53	B2	-.04	-.08
51	B2	-.01	-.09
54	A1	-.03	-.17
37	A1	.08	-.21
component weight (v_r)		.50	.09

Notes:

ACC = Ainsworth's classification category

? = unclassified

B3/4 = B3 or B4
C1 = first child component
C2 = second child component

ers therefore do not include it in their analyses (see e.g. Waters, 1978; Grossmann, Grossman, Huber, & Wartner, 1981). In *POA*, for instance, it is noted that for one-year olds distance interaction is a low-stress behaviour of low intensity, and that it differentiates less among the classification (sub) categories (p.246). Whether this is true for two-year olds is still a matter for investigation. We will come back to this point later. An acceptable label for the first scale component seems to be *intensity* of the reaction.

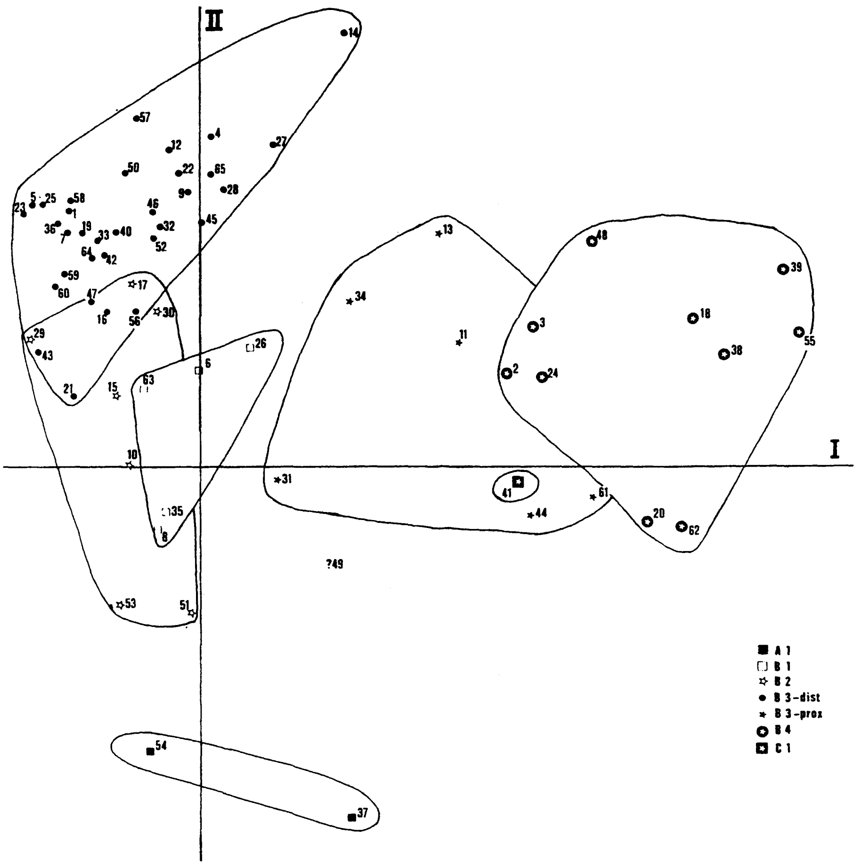


Fig. 8.2 *Attachment study: Child space (unscaled)*

The second component (S2) distinguishes between attachment behaviours, proximity seeking and contact maintaining, and behaviours antithetical to attachment, i.e. avoidance and resistance. It might be labelled as the *security seeking*. We will not discuss the third axis (S3) due to the small amount of variation explained by it (2%), even though it shows a theoretically important contrast between resistance and avoidance. It is, by the way, equally legitimate to define a PROX,CM-axis, and a RES,AVOI-axis by rotating the scale space.

Children. Table 8.3C and Fig. 8.2 show the two-dimensional child space for the 2x2x2-solution. The children have been labelled both by a sequence number and their Ainsworth classification subcategory (see Table 8.1C). These classifications are based on the same interactive scales as those in the present analysis. For the scoring, however, it is mainly the behaviour towards the mother which has been taken into account, instead of that towards both the mother and the stranger as in our analysis. The classification instructions are contained in *PAO* (p. 59-62; see also Swaan & Goossens, 1982), and require extensive training. One of the aims of applying three-mode principal component analysis to these data is to assess the adequacy of the scoring instructions. Psychological and medical research, for instance, have shown that people do not necessarily combine multivariate information in a very reliable way (see e.g. Sawyer, 1966; Linschoten, 1964, p.142ff.; Einhorn, 1972).

With respect to these data we will try to answer two questions. The first is whether the classification system is *consistent,* i.e. whether the children who occupy the same region in the child space, have the same Ainsworth classification . The second question is, whether the same scales to the same extent, are responsible for the grouping of the children, as specified in the scoring instructions. The grouping observed in our analysis may be the result of different combinations of scores. In other words, the present analysis is an attempt to validate the classification rules.

Ainsworth et al. (Ch. 6) applied discriminant analysis to check the adequacy of the classification system, but this involves the interactive scales twice: once to make the classification, and then to evaluate this classification by using the interactive scales as predictors in the discriminant functions. Here we use the interactive scales to group the children and to assess their contribution to this grouping simultaneously, and only after that we check the grouping against the classification. This provides a more adequate check of the appropriateness of the classification procedure.

The first impression from Fig. 8.2 is that on the whole a reasonable separation is possible between the B-subcategories, although on the basis of our analysis alone the divisions could not have been made. In addition, the two A1-children are in their proper places, as their score patterns on the interactive scales should be the mirror-image of the B3-children (see Table 8.1C). Furthermore, the one C1-child does not occupy a separate place. Finally, there are some B3-children seemingly belonging to the B4-children; they have been labelled 'B3-prox' for reasons to be discussed in section 8.7, where we will also try to provide the answers to the above questions. In the meantime we will use the Ainsworth classification to label the children, pretending we have already established its appropriateness.

8.4 INTERPRETATION OF THE CORE MATRICES

Explained variation. The core matrix indicates the relations between the various components of the three modes. For instance, the element c_{111} (=19.9) of the T3 core matrix (Table 8.4) indicates the strength of the relation between the first components of the three modes, and c_{221} (=13.5) the strength of the relation between the second components of the first and second modes in combination with the first of the third mode. As Table 8.4 shows, 30% of the SS(Total) is accounted for by the combination of the first components of the three modes, another 14% by c^2_{221}, and 3% each by c^2_{121}, and c^2_{211} (see section 6.9 for an explanation of this interpretation of the core matrix). We see that the differences between the children on the first component (C1) explain half of the fitted variation. This 50% can be partitioned as follows:

Table 8.4 *Attachment study: TUCKALS3 core matrix*

2x2x2-solution

(frontal planes)

child component (C1): B4 versus REST	components of interactive scales		proportion variation explained			
	S1 intensity of reaction	S2 security seeking				
components of episode:						
stress of situation E1	19.9	5.8	.30	.03	c_{111}	c_{121}
mother versus stranger E2	-5.8	13.5	.03	.14+	c_{211}	c_{221}
			V_1 = .50			
child component (C2): B3(dist) versus A1						
stress of situation E1	-6.7	3.0	.03	.01	c_{112}	c_{122}
mother versus stranger E2	-2.1	7.7	.00	.05+	c_{212}	c_{222}
			V_2 = .09			

3x3x3-solution

(frontal planes)

C1

	S1	S2	S3
E1	20.1	4.5	.6
E2	-4.8	13.7	-2.2
E3	-2.0	-2.3	-0.5

V_1 = .50

C2

	S1	S2	S3
E1	-6.5	2.8	2.6
E2	-2.0	6.8	0.1
E3	-7.1	-0.3	0.6

V_2 = .12

C3

	S1	S2	S3
E1	1.1	-6.7	-2.7
E2	-1.7	-0.8	0.0
E3	-4.9	-1.3	0.0

V_3 = .06

(a) due to c_{111} (30%): intensity of reaction (S1) due to the stress of situation (E1) for B4-children versus REST (C1);

(b) due to c_{221} (14%): security seeking (S2) with the mother versus stranger (E2) for B4-children versus REST (C1);

(c) due to c_{121} (3%): security seeking (S2) with stress of situation (E1) for B4-children versus REST (C1);

(d) due to c_{211} (3%): intensity of reaction (S1) with mother versus stranger (E2) for B4-children versus REST (C1);

The differences between the children on the second component (C2) contributes the remaining 9% explained variation, which can be broken down as follows

(e) due to c_{112} (3%): intensity of reaction (S1) due to the stress of the situation (E1) for B3-dist children versus A1-children (C2);

(f) due to c_{222} (5%): level of attachment (S2) with mother - stranger (E2) for B3-dist children versus A1-children (C2);

(g) due to c_{122} (1%): security seeking (S2) with stress of the situation (E1) for B3-dist children versus A1-children (C2).

Three-mode interactions. The percentages of explained variation only point to the important combinations, but do not indicate the direction of the relationship. This information can be found in the original (i.e. not-squared) core matrix. For the most important element of the core matrix the three-mode interaction between loadings on components is c_{111} (= +19.9). The plus sign indicates that

a. positive loadings on C1, S1 and E1 occur together:
   the more B4-like children are, the more intensely they react (= the higher above average their scores are on all scales except DI) in more stressful situations (= M5/S7 and M8);

b. negative loadings on C1 and S1 occur together with positive loadings on E1:
   the more negative a child loads on C1 the less intensely it reacts (= scores below average on all scales except DI) in more stressful situations (= M5/S7 and M8).

Or in slightly different terms:

° for B4-children (i.e. with positive loadings on C1) intensity of the reaction (S1) and stress of the situation (E1) are positively related;

° for children with negative loadings on C1 intensity of the reaction and stress of the situation are negatively related.

For the Goossens data the interpretation in terms of scores of idealized quantities (see section 6.9) is that an 'ideal' B4-child reacts intensely in stressful situations (c_{111} = 19.9), seeks much security with its mother-figure (c_{221} = 13.5), seeks moderate security in stressful situations (c_{121} = 5.8), reacts with moderately low intensity to the mother-figure (c_{211} = -5.8), and similarly for the other elements of the core matrix.

Extended core matrix. So far we have only looked at the interpretation of the core matrix of the Tucker3 model. As noted in section 6.9 the extended core matrix can be interpreted in essentially the same way as the TUCKALS3 core matrix in terms of the amount of explained variation.

We already noted the near equality of the components for the interactive scales and the episodes in the 2x2-solution and 2x2x2-solution in connection with Table 8.3A,B, consequently interpretations of those spaces are the same as before. The relationships between these components, as embodied in the frontal planes of the T2 core matrix, are given for a few selected children in Table 8.5. Four of the children were chosen because they are relatively close to one of the axes in the child space (i.e. 38, 57, 29, 37), and they can be considered 'idealized individuals' in the sense of e.g. Tucker & Messick (1963).

The frontal planes thus indicate how, for each child, the axes of the common space are related, just as was the case in the Tucker3 model for 'ideal' children. For instance, for child 38 (a B4-child) *intensity of reaction* (S1) and *stress of the situation* (E1) are positively related (see Table 8.5), as are *security seeking* (S2) and the *mother versus stranger* distinction (E2), while the other combinations are immaterial. For child 35, by comparison (a B1-child), none of the relationships seem very relevant (see section 8.7 for a discussion of this phenomenon). Note also that the two A1-children (37 and 54) have very different patterns of relationships, notwithstanding their similar position in the child space (Fig. 8.2).

Roughly one can conclude that children on the first child dimension (C1) weight the *intensity* (E1) - *stress* (S1) combination

Table 8.5 *Attachment study: TUCKALS2 core planes for selected children*

	B4 (38)		B3 (57)		B2 (29)		B1 (35)	
	S1	S2	S1	S2	S1	S2	S1	S2
E1	5.7	0.9	-2.1	-0.7	-2.4	-0.2	-0.2	-1.0
E2	-0.5	5.1	-0.7	1.4	0.4	-0.6	1.1	0.1
*)	.30	.07	-.04	.21	-.09	.08	-.02	-.02

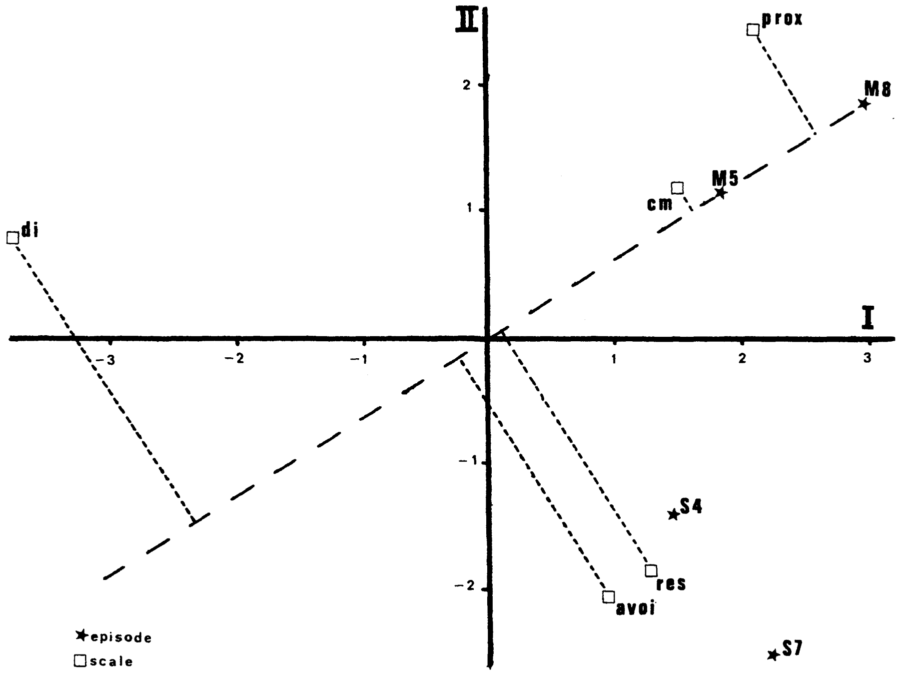
	A1 (54)		A1 (37)		C1 (41)		B2 (51)	
	S1	S2	S1	S2	S1	S2	S1	S2
E1	0.6	-2.2	3.0	-3.2	3.7	-.07	0.5	-0.0
E2	2.3	-0.8	-0.1	0.4	-1.0	2.9	0.2	-0.9
*)	-0.3	-.17	.08	-.21	.18	-.01	-.01	-.09

*) T3 component loadings (See Table 8.3C)

Notes: B4 (38): child nr. 38 - Ainsworth classification category B4
S1 (S2): first (second) scale component
E1 (E2): first (second) episode component

and the ***mother versus stranger*** (E2) - ***security seeking*** (S2) combination with a ratio similar to the ratio of c_{111} to c_{221} in the T3 analysis, and that the overall size of the elements determines their position on the C1 component: high positive numbers on the diagonal of the TUCKALS2 core plane (e.g. for child 41 and child 38) lead to highly positive loadings on C1, and moderately negative numbers (e.g. for child 29) lead to moderately negative loadings. On the negative side of the second child component (C2) there are children who emphasize the (E1, S1) combination, but not or hardly the (E2, S2) combination (child 37), and on the positive side of C2 (child 57) the situation is reversed, i.e. (E2, S2) is high and (E1, S1) low. This distinction corresponds with the opposite signs in the second frontal plane of the T3 analysis.

A 'IDEAL' B4 CHILD



B 'IDEAL' B3-DIST CHILD

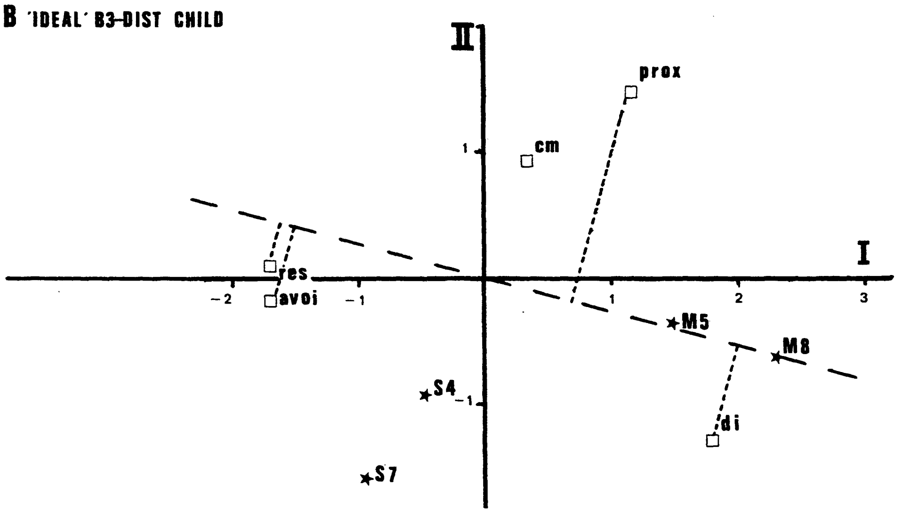


Fig. 8.3 A,B *Attachment study: Joint plots of episodes and interactive scales.*

8.5 JOINT PLOTS

With joint plots (see section 6.10) we can examine in some detail the relationships between the interactive scales and the episodes for each ideal-type child or child component. In Fig. 8.3A,B we present the joint plots for the two child components. The following characterization for the children loading on the positive side of the first component C1, i.e. B4-children, can now be made:

(a) they have high scores on proximity seeking and contact maintaining towards the mother (in episodes M5, M8), and they score about twice as high in M8 as in M5. With a high score we mean relatively to the overall scale means, as we have removed these means for all interactive scales.

(b) they have high scores on resistance and avoidance towards the stranger (in S4 and S7), and nearly twice as high in S7 as in S4.

(c) they show roughly average resistant and avoidant behaviour towards the mother in M5 and M8, even somewhat below average on avoidance. Similarly, proximity seeking and contact maintaining towards the stranger have average values.

(d) the scores on distance interaction do not discriminate between the mother and the stranger, and they are below average. There is less distance interaction in the later episodes.

These interpretations are derived from the fact that the scales can be seen as points and the episodes as vectors or directions in the common space, or vice versa. In this case the former approach is to be preferred because the episodes are fixed, i.e. they are elements of the design. The relative importance of the various scales at any episode can then be assessed from their perpendicular projections on the vectors as is shown for M5 and M8 combined.

For the positive scores on the second child component, i.e. the B3-dist children, the characterization is (see Fig. 8.3B):

(a) low scores on resistance and avoidance towards the mother, coupled with average contact maintaining and proximity seeking. High distance interaction increasing further in M8.

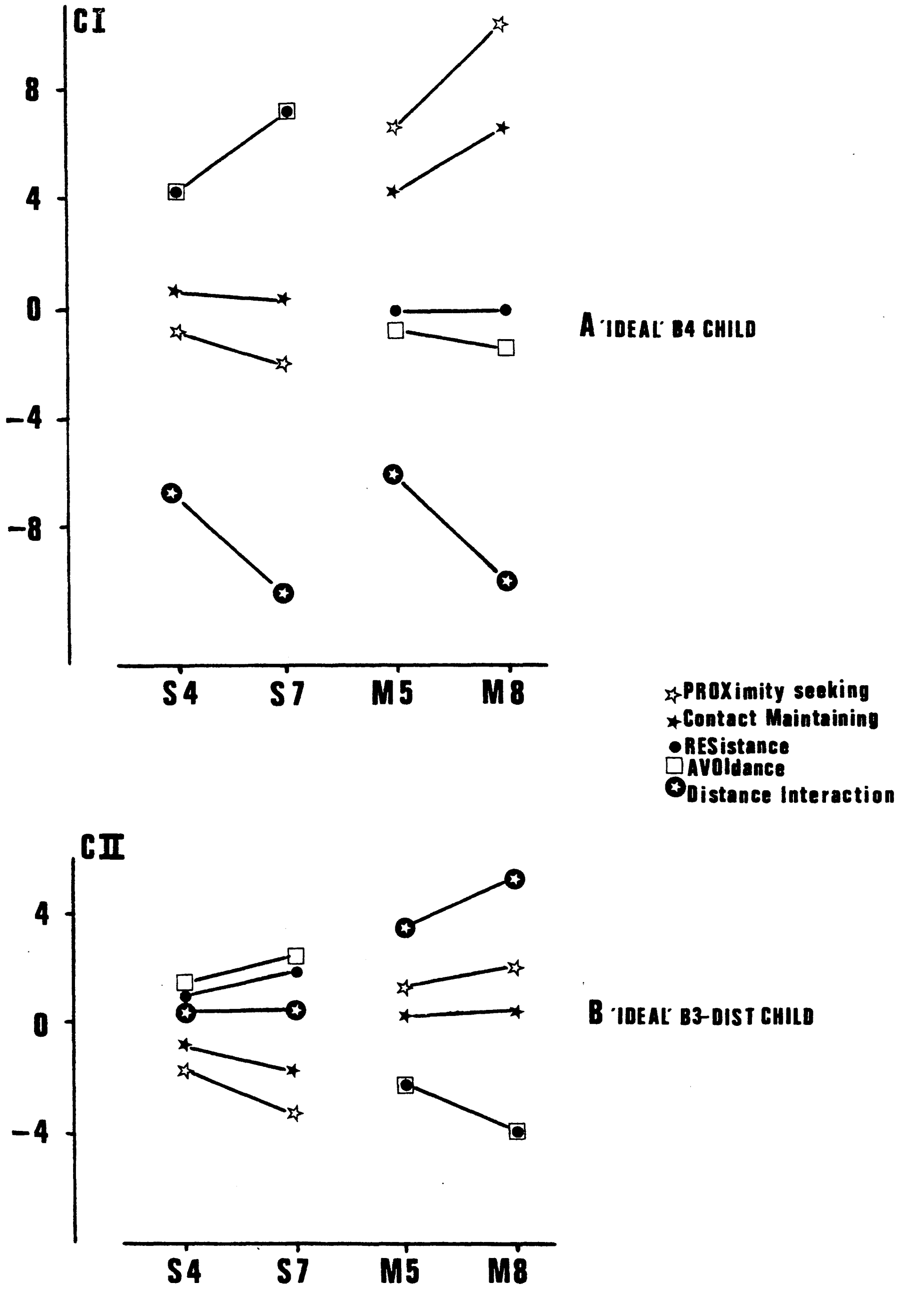


Fig. 8.4 *Attachment study: Component scores for episode-scale combinations.*

(b) low scores on proximity seeking and contact maintaining towards the stranger, with lower scores on proximity seeking. Average resistance, avoidance, and distance interaction with a slight increase in the avoidance measures in S7.

For 37, an Al-child, the mirror image of the above observations is true as he/she lies on the negative side of the second child component (C2). These relationships are displayed in Fig. 8.4, where the component scores are given for the 'ideal' B4-child, and the 'ideal' B3-dist child.

8.6 FIT OF THE SCALES, EPISODES, AND CHILDREN

In Table 8.6 the sums of squares for the scales and episodes are shown. From the SS(Total)s for episodes we see that the variability as expressed by the sums of squares increases with the later episodes, as children deviate more from the scale means, or probably show more variation among themselves. With respect to the scales we see that *contact maintaining* has relatively little variability, while *distance interaction* has considerably more. From the residual sums of squares we note that the scales fit more or less equally well, irrespective of their total sum of squares, but that the configurations derived and discussed above are for a large part determined by the last two episodes. The structure described is, therefore, more representative of the later behaviours than the earlier ones. This explains, for instance, why an added third episode component shows an early versus late character; primarily the earlier episodes will then be fitted better.

Fig. 8.5 is the sums-of-squares plot which shows the residual sums of squares versus the fitted sums of squares for the children from the 2x2x2-solution.

A number of features are particularly noteworthy. The B4-children fit well, have large sums of squares, and dominate the solution. Furthermore, there is a large group of B3-children (mainly B3-dist) which have small total sums of squares (thus they score about average on all scales), and most of their variation is fitted well. On the other hand, none of the Bl- and B2-children fit very

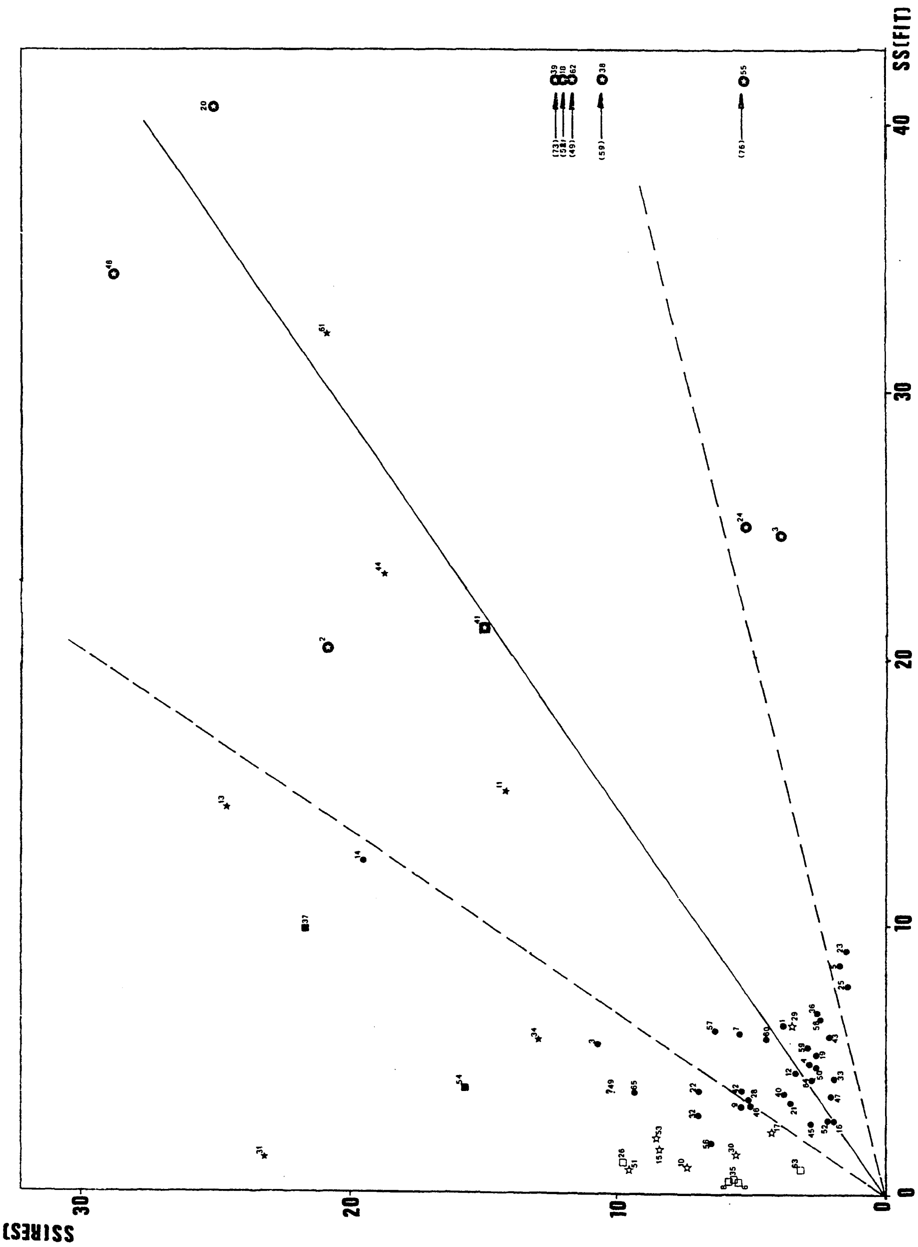


Fig. 8.5 *Attachment study: Sums-of-squares plot for children.*

Table 8.6 *Attachment study: Sums of squares*

A. *Episodes* (mode 1)

		2x2x2-solution				*3x3x3-solution*	
epi-sode	SS(Total) stand.	SS(Fit) stand.	rel.	SS(Res) stand.	rel.	SS(Res) stand.	rel.
S4	.16	.07	.40	.10	.60	.10	.59
M5	.21	.09	.44	.12	.56	.06	.27
S7	.29	.18	.63	.11	.37	.09	.30
M8	.33	.24	.74	.09	.26	.08	.23
over-all	1.00	.59		.41		.32	

B. *Interactive scales* (mode 2)

		2x2x2-solution				*3x3x3-solution*	
scale	SS(Total) stand.	SS(Fit) stand.	rel.	SS(Res) stand.	rel.	SS(Res) stand.	rel.
PROX	.23	.14	.61	.09	.39	.06	.27
CM	.10	.05	.54	.05	.46	.04	.41
RES	.15	.08	.52	.07	.48	.06	.41
AVOI	.17	.08	.44	.09	.56	.08	.46
DI	.35	.24	.68	.11	.32	.07	.21
over-all	1.00	.59		.41		.32	

Notes: stand. = standardized or divided by the overall SS(Total).
rel. = relative sum of squares, which is defined as:

$$\text{relative SS (Res) of episode S4} = \frac{\text{SS(Residual) of episode S4}}{\text{SS (Total) of episode S4}}$$

well into the overall pattern, but we have to remember that there are only few of them. Their total sums of squares are not very large, but their relative residual sums of squares are. Finally, there is a number of children which couple considerable sums of squares with little fit, indicative of either another organization of the scale and episode relationships, or large amounts of random variation. In fact, the two A1-children (37 and 54) belong to this group.

8.7 DISCUSSION

Keeping in mind the provisional character of the data, there are some conclusions that can be drawn with respect to the example. In the first place, we note that three-mode principle component analysis has succeeded in showing individual differences between the children, and characterizing the kind and degree of these differences. Furthermore, the analysis presented here supports to a large degree the *consistency* of the classification procedures as described by Ainsworth et al. in *POA*, especially for the B-children. The consistency follows from the grouping of children belonging to the same category. The presence of only two A-children and a single C-child precludes any serious statements about these classification categories, apart from the observation that their position in the child space (Fig. 8.2) agrees with what one would expect.

In section 8.3 we noted the presence of two groups of B3-children. In Fig. 8.2 they were labelled *B3-prox* and *B3-dist*. The classification instructions in *POA* (p. 61) for B3-children (see also Swaan & Goossens, 1982) also suggest that there are two types of B3-children: those who actively seek physical contact with their mothers (B3-prox), and those who seem especially 'secure' in their relationship with their mother, and are content with mere interaction from a distance with and proximity to the mother without seeking to be held (B3-dist). It is possibly due to the greater ability of communicating at a distance on the part of two-year olds that there are more children in the B3-dist than in the B3-prox group in Goossens' sample. For one-year olds the reverse seems to be true (see Goossens, Note 1, for further details).

In Table 8.7 the characterizations of the children (derived from Fig. 8.4), occupying the extremes of the axes in Fig. 8.2 (child space) are presented. Comparing this table with Table 8.1C (reproduced in part here) shows global agreement and disagreement in detail. The most conspicuous differences are related to resistance and distance interaction. The comparison for resistance is probably biased by the absence of extremely resistant (C)-children, and 'high resistance' in Goossens' sample might be average when compared to the resistant behaviour of C-children. The differences between distal behaviours are, most likely related to the age differences.

Table 8.7 *Attachment study: Comparison of Ainsworth's and TUCKALS classifications*

	AINSWORTH						TUCKALS				
	PROX	CM	RES	AVOI	DI		PROX	CM	RES	AVOI	DI
A1	-	-	-	++	-	A1	o	o	H	H	L
						B3-dist	o	o	L	L	H
B3-prox	++	++	-	-	-	B3-prox	H	H	o	o	L
B4	++	++	(+)	-	-	B4	HH	HH	o	o	LL

\- low +(+) moderate to high
(+) low to moderate
\+ moderate ++ high

LL = low H = average to high
L = low to average
o = average HH = high

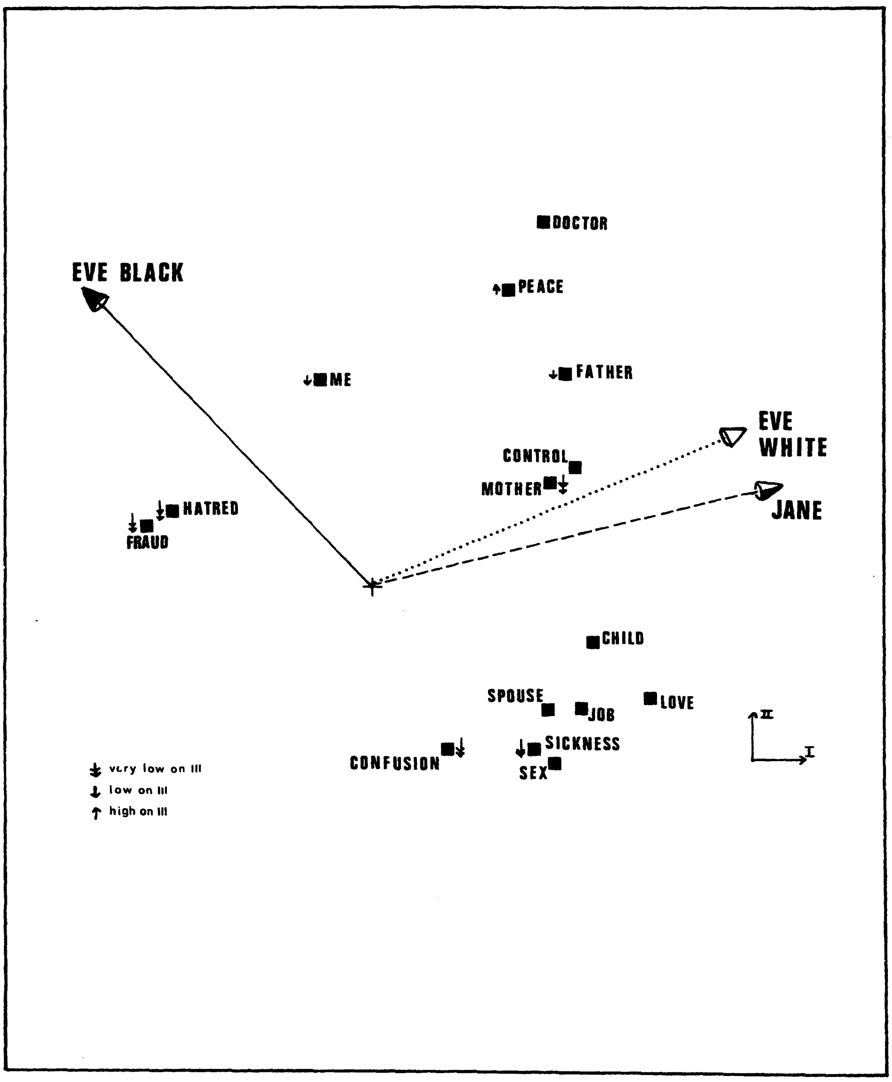
A number of problems remain. One is the low number of A-children compared to the number found in samples of one-year old children. One of the explanations might be that this is due to less avoidant behaviour of two-year old children. Another, by now more likely explanation is that it is due to a somewhat non-standard scoring procedure for avoidance (see Goossens, Note 1).

A further possible problem are the ill-fitting B1- and B2-children. Two reasons might be put forward in this respect. One is that they have approximately average scores on all scales so that we are trying to fit their individual error, rather than any meaningful variation; otherwise it might be that their way of reacting to the strange situation cannot be fitted very well together with the other children. Their small number might preclude finding a separate dimension for themselves. Clearly these conjectures could and will be further investigated.

SEMANTIC DIFFERENTIAL DATA

9

triple personality study



9.1 INTRODUCTION

In this section we will present an example of the power of three-mode principal component analysis in constructing one unified description of data collected under different circumstances, and (possibly) referring to the same underlying structure. Specifically, we will analyse data from probably the most famous case of a multiple personality: Eve White, Eve Black, and Jane (Thigpen & Cleckley, 1954). Osgood & Luria (1954)[1] published scores on semantic differential scales for each personality at two occasions (testings I and II). In essence the data set is a four-mode one, i.e. personality × testing × concept × scale. We will, however, treat them as three-mode data, and the *6* administrations (k-mode) of *10* scales (j-mode) by *15* concepts (i-mode) are the material on which this re-analysis is based. Besides presenting a unified analysis of these data, the example sets out to show how individual differences in the use of semantic differential scales can be analysed with three-mode principal component analysis. Other examples of three-mode analysis on semantic differential data can be found in the *references* , via the subject classification of applications in the Appendix (see also Kroonenberg, 1983, in press).

9.2 THE SEMANTIC DIFFERENTIAL TECHNIQUE2

The semantic differential technique is a combination of association and scaling procedures designed to give an objective mea-

---

1. Page references are to the reprinted version in Snider & Osgood (1969).
2. This subsection is at times an almost literal citation from Osgood & Luria (p.505).

sure of connotative meaning. A linguistically complex assertion such as "My father has always been a rather submissive person", can be at least partially represented on bi-polar seven point scales

MY FATHER active – : – : – : – : $\overset{X}{-}$: – : – passive

MY FATHER soft – : $\overset{X}{-}$: – : – : – : – : – hard

The greater the strength of association, e.g. "extremely submissive, a regular doormat", the more polarized towards 1 or 7 the check mark on the scales. Since many scales of judgement are highly intercorrelated (e.g. good-bad, fair-unfair, honest-dishonest, kind-cruel, and so forth, all reflect mainly the single "evaluative" factor in judgements), a limited number of dimensions can be used to define a semantic space within which the connotative measuring of any concept can be specified. Factor analytic studies of semantic differential data consistently show that there are three major dimensions of rating response: *Evaluation, Activity,* and *Potency* (see e.g. Heise, 1969, p. 412-415).

Table 9.1 *Triple personality study: concepts and scales*

Concepts *

	LOVE	mental SICKNESS	self-CONTROL
	CHILD	my MOTHER	HATRED
my	DOCTOR	PEACE of mind	my FATHER
	ME	FRAUD	CONFUSION
my	JOB	my SPOUSE	SEX

Scales **

valuable - worthless	E	fast - slow	A	large - small	P,a
clean - dirty	E	active - passive	A	strong - weak	P
tasty - distasteful	E	hot - cold	A	deep - shallow	P,e
		relaxed - tense	E,a		

* The abbreviations used for the concepts are in upper case

** E = evaluation, A = activity, P = potency; upper case letters indicate high loadings on a factor (in 'standard' settings); lower case letters indicate medium loadings on a factor (in 'standard' settings).

Source: Osgood & Luria (1954; 1969, p. 506).

The form of semantic differential used in the study of the triple personality of Eve White, Eve Black, and Jane is given in Table 9.1 (adapted from Osgood & Luria, p.506).

9.3 OSGOOD & LURIA'S ANALYSIS

Osgood & Luria, lacking a technique for simultaneously treating all their data, obtained measures of semantic similarity and structure by computing generalized distances between each pair of concepts for each of the six administrations of the scales. This generalized distance $d_{ii'k}$ between concept i and i' for adminstration k summed over the m(=10) scales is defined as

$$d_{ii'k} = \sum_{j=1}^{m} (z_{ijk} - z_{i'jk})^2 \qquad i,i' = 1,\ldots,\ell \;; k=1, \ldots, n$$

For the justification of this measure they refer to Osgood & Suci (1952; see, however, Torgerson, 1958, pp. 294-296). The $d_{ii'k}$ used by Osgood & Luria were derived from factor scores which were the result of factoring the concept by scales matrices for each person (omitting relaxed-tense), rather than the raw data. As they used only three factors their $D_k = \{d_{ii'k}\}$ matrix has rank 3 as well, and it can, therefore, be plotted 'error-free' in three dimensions which are found by factoring the D_k matrices (see Osgood & Suci, 1952). These three factors were computed by Osgood & Luria for all six administrations, and were presented as 'ball-diagrammes' (see Figure 9.3).

In order to compare the six administrations the intercorrelations of the D_k were computed. The concept distances between testings (t,t') for each personality p (p=1,2,3) over all scales, i.e.

$$d_{iptt'} = \sum_{j=1}^{m} (z_{ijpt} - z_{ijpt'})^2 ,$$

and between personalities (p,p') for each testing t

$$d_{ipp't} = \sum_{j=1}^{m} (z_{ijpt} - z_{ijp't})^2$$

were also computed, to assess the differences between and within personalities. These concept distances are given in Table 9.2. From this table we see that the largest differences are between Eve Black on the one hand and Eve White and Jane on the other. The differences between Eve White and Jane are so small that concepts singled out by Osgood & Luria as most discriminating between them have concept distances of the same order of magnitude as the test-retest concept distances, except for sex.

Table 9.2 *Triple personality study: concept differences between personalities and testings*

	Between personalities Within testings						Within personalities Between testings		
	D_{W-B}		D_{J-B}		D_{W-J}		D_{I-II}		
Concepts	I	II	I	II	I	II	B	W	J
Child	1.65	1.40	1.47	1.41	.68	.54	.96	.71	.37
Love	1.58	1.44	1.62	1.81	.35	.57	.67	.42	.23
Hatred	1.54	1.31	1.37	1.19	.51	.23	.19	.51	.44
Fraud	1.46	1.35	1.29	1.22	.73	.34	.12	.64	.40
Job	1.30	1.43	1.19	1.54	.49	.43	.62	.27	.42
Sickness	1.24	1.38	1.30	1.47	.40	.32	.45	.45	.19
Me	1.21	1.40	.83	.77	*.60*	*.88*	.32	.36	.42
Sex	1.10	.62	1.45	1.76	*.63*	*1.20*	.34	.62	.47
Father	1.06	.60	1.06	.43	.25	.43	.71	.53	.09
Confusion	.86	.96	.98	.88	.71	.42	.67	.64	.44
Peace	.86	.66	.81	.61	.21	.28	.35	.40	.41
Control	.78	.80	.92	1.01	.34	.39	.25	.32	.24
Mother	.71	.78	1.02	.68	.66	.23	.78	.54	.46
Spouse	.67	.96	1.04	1.75	*.61*	*.89*	.62	.30	.47
Doctor	.23	.23	.30	.12	.28	.25	.05	.15	.27

Note: numbers in italics indicate the concepts that (according to Osgood & Luria) serve best to characterize differences between Eve White and Jane

Adapted from Osgood & Luria (1954; 1969, p.513).

The most conspicuous aspect of Osgood & Luria's analysis is its indirect way in arriving at a geometric representation of the concepts. Furthermore, no goodness-of-fit is reported for their three-dimensional solution of concepts, nor is information given on the acceptability of a three-dimensional solution for the scales.

In a later discussion of this paper Osgood, Suci & Tannenbaum (1957) present the three (rotated) factors for the scale spaces of the first testing (I) of Eve White, Eve Black, and Jane. These factor loadings show a strong first rotated factor (49%, 59%, and 48% explained variation for the personalities respectively) on which nearly all scales load positively and which is interpreted as a 'general' evaluative factor. The second and third factors resemble each other far less (as shown by their Spearman correlations .56, .14, and .59 for the second factors, and .87, .24, and .21 for the third factor respectively), but Osgood et al. see sufficient similarities in them to state "we have evidence, then, for essentially the same three major factors operating in the several personalities of this disturbed patient, although there is considerable shifting in meanings of specific scales between personalities ..." (p. 262). Inspecting their factor loadings and the correlations between them we tend to think they overstate their case. In addition, it is questionable how useful the statement about shifting scales is without reference to the concepts to which the dimensions and scales apply. Because of this it should be useful to look at the data in their entirety using a less arbitrary method for their analysis, and attempt to answer the question: "In which way do the three personalities differ, and in which way do they resemble each other ?"

9.4 PREPROCESSING OF THE DATA

Before analysis the scales valuable-worthless, tasty-distasteful, deep-shallow, active-passive, were recoded, so that they also were scored in a 'positive' manner analoguous to the presentation in Table 36 in Osgood et al. (1957).

As in our other examples, a central question is the treatment of the means before the analysis proper. As it is assumed in semantic differential research that the centre of the scale (here :4) is the neutral point, and that a concept which has a 4 on all scales is a "meaningless" concept (cf. Osgood & Luria, p. 507), it seems most proper to subtract the scale midpoint 4 from all values. An alternative would be to subtract the scale averages at each administration, as was probably done during the factor analysis performed by Osgood & Luria. The disadvantage of the latter approach would have been that shifts in overall level of scoring between personalities and testings would have been eliminated from the analysis.

9.5 THREE-MODE ANALYSIS

After investigating solutions with varying numbers of components for the three modes, it was decided to report the details of the *3x2x2-solution*, i.e. 3 concept components, 2 scale components, and 2 personality components. In passing we will refer to components from other solutions.

Scale space. In contrast to most analyses of semantic differential data only *two* scale components were deemed necessary. With these two components most similarities and differences between the personalities can be described. A third scale component explains, by the way, another 3% of the total variation; this contrasts fast versus large, clean, and valuable. Table 9.3 and Fig. 9.1 show the two-dimensional scale space. Conspicuous is the ab-

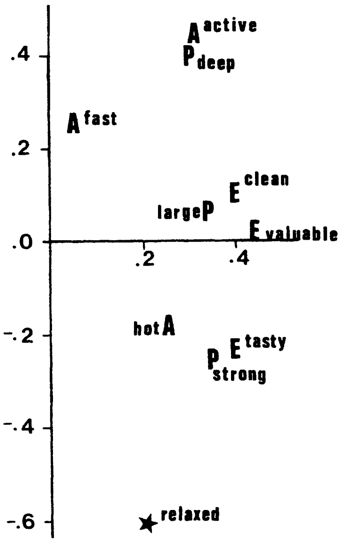


Fig. 9.1 *Triple personality study: scale space*

Table 9.3: *Triple personality study: scale space*

scale		S1	S2
valuable	E	42	1
clean	E	39	12
tasty	E	38	-23
fast	A	5	25
active	A	31	45
hot	A	25	-18
large	P	34	8
strong	P	33	-26
deep	P	30	42
relaxed	E,a	21	-62
% explained variation		59	11

Note: decimal points omitted

sence of an EPA-structure. It seems a matter of taste what to call the axes, clearly the standard terminology is only partly helpful. Osgood et al. (1957) labelled their first (rotated) axes, which resemble somewhat ours -S1, *evaluation*. They referred to the second axes of Eve White I, Eve Black I, and Jane I as *potency* axes. The differences between their second axes, and their differences with ours preclude such a label in this case. Our second component (S2) is dominated by relaxed (E,a), active (A), and deep (P), and seems difficult to interpret within the standard framework. We will come back to this later in connection with the discussions about concepts and personalities.

Concept space. In comparison with Osgood & Luria's complicated way to derive the concept space, the configuration of concepts emerges naturally in three-mode analysis, and its dimensionality can be assessed more or less independently of the dimensionality of the scale space. Three dimensions were necessary to give a reasonable representation of the concept space (Table 9.4)

Table 9.4: *Triple personality study: concept space*

	C1	C2	C3
doctor	23	53	- 1
peace	19	45	20
father	26	31	-21
control	28	19	1
mother	24	17	-32
child	31	- 6	-13
love	39	-14	5
job	30	-16	- 8
spouse	25	-17	8
sickness	23	-23	-38
sex	26	-25	5
confusion	11	-23	-50
me	- 8	30	-24
hatred	-27	11	-43
fraud	-30	8	-38
% explained variation	38	21	10

Note: decimal points omitted

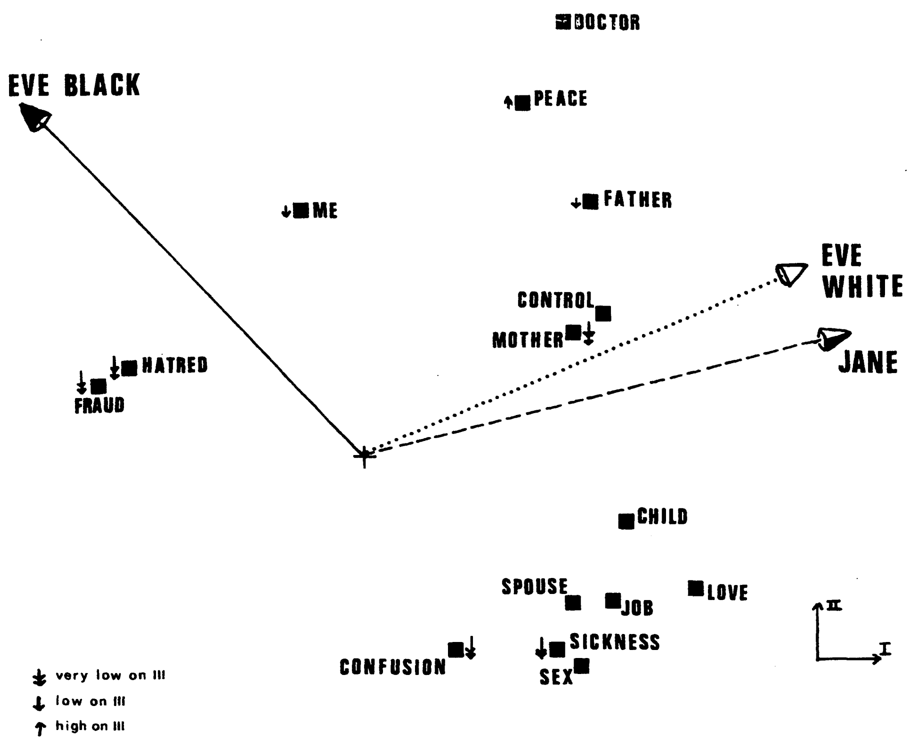


Fig. 9.2 *Triple personality study: Concept space*

It is very instructive to compare the first two dimensions of the concept space (Fig. 9.2) with the ball-diagrammes of Osgood & Luria reproduced here in Fig. 9.3. The arrows in Fig. 9.2 roughly correspond to the longest axis in the concept spaces for each of the personalities in the Osgood & Luria analysis (Fig.9.3). The TUCKALS concept space thus represents the characteristics of all three personalities simultaneously. The large differences between Eve White and Jane on the one hand, and Eve Black on the other hand are also evident in the *personality space* (Table 9.5), the *TUCKALS3-core matrix* (Table 9.6), and the *TUCKALS2-core matrix* from a 3x2-solution (Table 9.7). These tables give a coherent, quantitative indication of the differences between the personalities at different levels of summarization, in contrast with Osgood & Luria, who

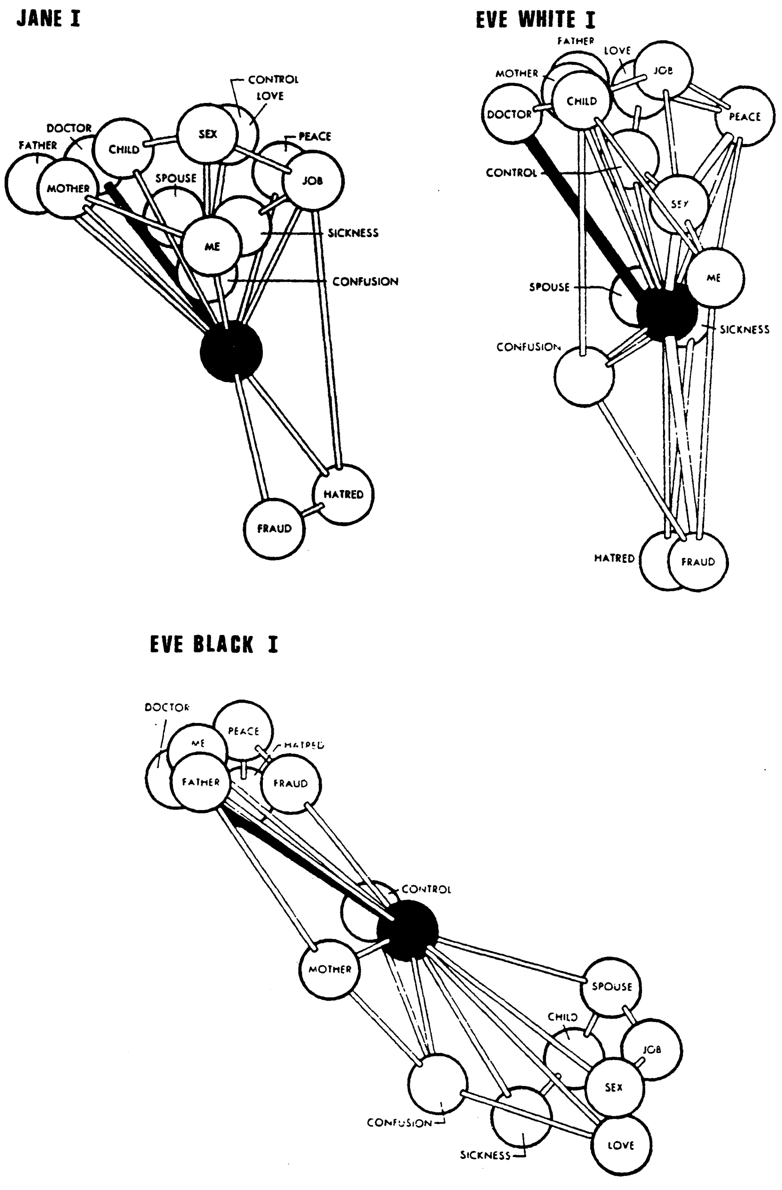


Fig. 9.3 *Triple personality study: Osgood & Luria's concept spaces*

interpret largely on the basis of qualitative comparisons between the personalities. It is especially the compactness of representation and the simultaneous portraying of the relationships between administrations which makes a three-mode analysis attractive in this case.

Table 9.5: *Triple personality study: personality space*

		P1	P2
Eve Black	I	-11	62
Eve Black	II	-13	76
Eve White	I	47	10
Eve White	II	44	16
Jane	I	51	6
Jane	II	54	3
% explained variation		45	24

Table 9.6.: *Triple personality study: TUCKALS3 core matrix*

		personality component 1 (Eve White & Jane)		personality component 2 (Eve Black)	
		S1	S2	S1	S2
concept components	C1	18	0	- 4	2
	C2	3	- 1	13	1
	C3	0	- 9	- 0	- 4

% explained variation by combinations of components

concept components	C1	35	0	2	0
	C2	1	0	20	0
	C3	0	8	0	1

Table 9.7.: *Triple personality study: TUCKALS2 core matrix*

Testings		Eve Black		Eve White		Jane	
		S1	S2	S1	S2	S1	S2
I	C1	- 5	0	8	1	9	0
	C2	8	0	3	- 1	2	0
	C3	- 1	- 1	2	- 5	- 1	- 5
II	C1	- 6	2	7	0	10	0
	C2	10	1	4	- 1	3	0
	C3	0	2	0	- 5	- 1	- 4

Concept-scale interactions. In this subsection we will turn to a description of the relationships between the scales and concepts for each of the personalities. For the moment we will treat Eve White and Jane as one personality, and only comment on their differences later on.

Eve Black's scale and concept relationships are given by a joint plot in Fig. 9.4, and they are summarized in Table 9.7 by her T2-core planes. The location of the concepts is a good compromise of the figures for Eve Black I & II as presented by Osgood & Luria. Insight into the scale-concept relationships is especially important in this case as the scale space does not show the usual EPA-structure, so that these labels are not applicable here. Summarizing the relationships in a few words one could say that all concepts related to day-to-day life (job, spouse, child, sex, love) are evaluated negatively, and are considered neutral with respect to scales as active, deep, and relaxed. Those concepts related to Eve Black's mental make-up (confusion and mental sickness) are also evaluated negatively, but somewhat active and deep and rather tense as well. Eve Black regards with favour her therapist, herself, peace of mind, hatred, and fraud, and has a moderately favourable opinion of her parents, as well as a moderately active and deep, and a rather tense judgement of them.

From Tables 9.2 and 9.5 it is clear that *Eve White* and *Jane* are very much alike, as illustrated in Fig. 9.5 and reasonably 'normal'. All concepts related to day- to-day life and therapy are positively evaluated, while hatred and fraud are not. Me is seen as a neither good nor bad concept and somewhat fast, weak and distasteful, as well as rather tense, active and deep. Noteworthy is furthermore that confusion and sickness are neutrally evaluated, and are very tense, active, weak, distasteful and cold.

The differences between Eve White and Jane, as perceived by Osgood & Luria, do not show very clearly here. It is possible to derive a third personality component which contrasts Eve White and Jane, but it explains less than 1% of the total variation. An analysis of the concept-scale relationships for this third personality component shows that Eve White compared to Jane finds confusion and sickness more distasteful, weak, and tense. In other

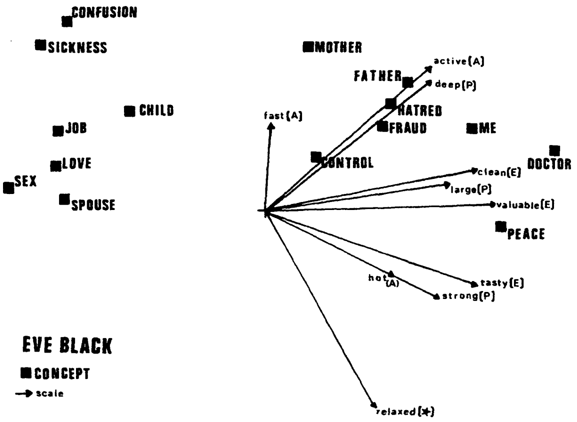


Fig. 9.4 *Triple personality study: Eve Black's concept-scale space*

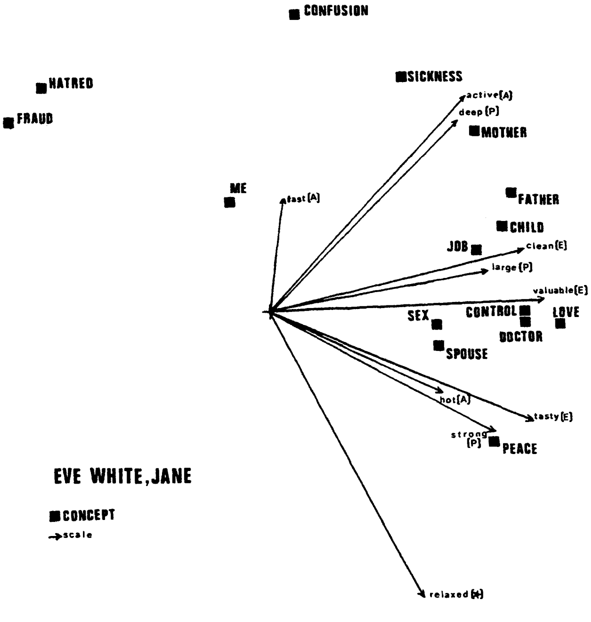


Fig. 9.5 *Triple personality study: Eve White's and Jane's concept-scale space*

words, Jane apparently thinks less unfavourable of the concepts related to her mental state. The effect is, however, rather small. These differences can be discerned, by the way, in Osgood & Luria's ball-diagrammes, but the different reactions to these concepts do not show up in Table 9.2, probably due to summing over scales.

9.6 DIFFERENCES WITH OSGOOD & LURIA

There are differences between our analyses, and those of Osgood & Luria. In the first place we find a far stronger similarity between Jane and Eve White than Osgood & Luria suggest, and the differences we do find are not those they mention as important.

Their conclusion that "Jane is becoming *less* diversified semantically (more 'simple-minded') rather than the reverse" (p. 516), with "... all of her judgments tending to fall along a single factor of *good-strong* vs. *bad-weak*" (p. 514), is only very weakly supported by our analysis. If we take 'simple-mindedness' to mean that one of the combinations of scale and concept axes increases at the cost of the others, then indeed we observe from Jane's T2-core planes (Table 9.7) that $\tilde{c}_{C1,S1}$ increases from 9 to 10, and the other large element $\tilde{c}_{C3,S2}$ decreases (in absolute size) from -5 to -4. Thus the evaluative-like first scale component becomes more important with respect to day-to-day concepts, therapy, and hatred and fraud, while active, deep, tense judgements of mother, sickness, confusion, hatred and fraud become less. If this change warrants the strong statement of Osgood & Luria is rather doubtful.

The statement that there is an "increasing simplification in structure characteristic of all three personalities" (p. 517) cannot be supported in the same manner (see Table 9.7). A more detailed analysis, and possibly a replication of their analysis coupled with separate analyses via singular value decomposition for each of the personalities might show how these differences arise - a course which we will not pursue here.

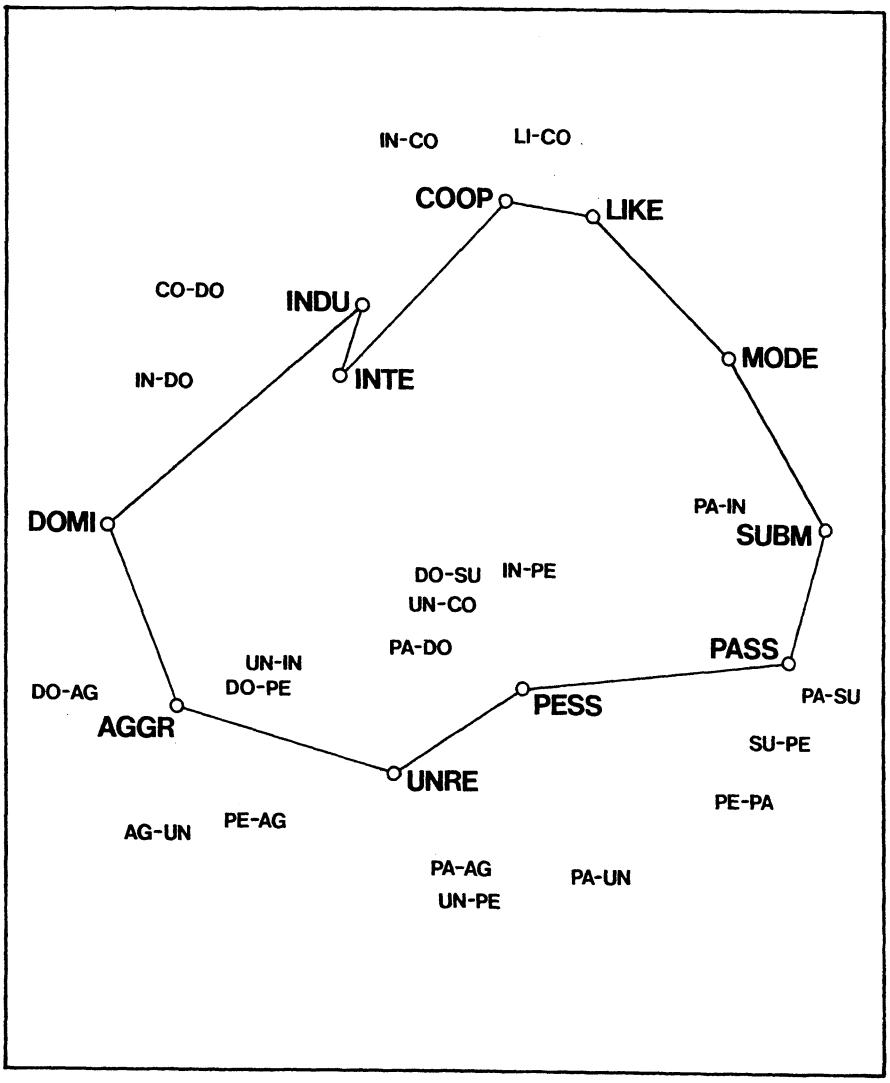
9.7 CONCLUDING REMARKS

In this example we showed how three-mode principal component analysis can fruitfully be used to provide a unified description of the scale and concept usage in the case of a multiple personality. At the same time this study can serve as an example of how individual differences can be handled in semantic differential research far more easy than was customary in the early stages of its development.

ASYMMETRIC SIMILARITY DATA

10

ITP study



10.1 INTRODUCTION

The main purpose of this section is to show how theoretical subjects may serve to aid the interpretation of subject spaces. A *theoretical subject* is defined as the collection of scores (or response pattern) on the variables in a study which has been derived on the basis of theoretical considerations. In other words, substantive knowledge is used to derive how a subject should score if he conformed to (a particular aspect of) the theory. The particular illustration used here is taken from Van der Kloot & Kroonenberg (1982).

10.2 THEORY, DESIGN, AND DATA

Theory. The notion that people use naive, common sense, or *implicit theories of personality (ITP's)* when they form impressions of another person's personality was introduced in 1954 when Bruner and Tagiuri proposed a cognitive approach to the study of person perception. In its original meaning, a person's ITP is a set of perceived or expected relations among personality traits; these perceptions and expectations may vary from person to person. Since then, numerous studies have been conducted on various aspects of the concept of the ITP. These studies were reviewed by Schneider (1973).

Van der Kloot & Van den Boogaard (1978) conducted an experiment to gain insight in the way people process information about other persons. Experiments in this field usually follow Asch's (1946) original paradigm in which a stimulus person is described by a number of personality trait adjectives. The subjects have to express their impressions of a stimulus person by either checking similar adjectives in a checklist, or by giving numerical judgments on one or more rating scales.

Design and data. Van der Kloot and Van den Boogaard used 11 personality trait adjectives: *Likeable, cooperative, intelligent, industrious, dominant, aggressive, unreliable, pessimistic, passive, submissive, and modest.* These traits were selected because earlier research had shown that these stimuli lie on a circle in the order in which they are presented above. These stimuli were used in two experimental tasks.

In the first task, subjects had to rate 11 stimulus persons. Each stimulus person was described by one of the adjectives mentioned above (for instance: somebody is *aggressive*). In the second task the subjects had to rate 20 stimulus persons, each described by combinations of two personality trait adjectives (see Table 10.1). For further details see Van der Kloot & Kroonenberg (1982).

Table 10.1 *Combinations of adjectives in experimental task*

Likeable-cooperative	(LI-CO)	Unreliable-intelligent	(UN-IN)
Cooperative-dominant	(CO-DO)	Unreliable-pessimistic	(UN-PE)
Intelligent-cooperative	(IN-CO)	Pessimistic-aggressive	(PE-AG)
Intelligent-dominant	(IN-DO)	Pessimistic-passive	(PE-PA)
Intelligent-pessimistic	(IN-PE)	Passive-intelligent	(PA-IN)
Dominant-aggressive	(DO-AG)	Passive-dominant	(PA-DO)
Dominant-pessimistic	(DO-PE)	Passive-aggressive	(PA-AG)
Dominant-submissive	(DO-SU)	Passive-unreliable	(PA-UN)
Aggressive-unreliable	(AG-UN)	Passive-submissive	(PA-SU)
Unreliable-cooperative	(UN-CO)	Submissive-pessimistic	(SU-PE)

The descriptions of the stimulus persons were presented in two booklets, each preceded by an instruction page. The descriptions were printed on top of each page, and were followed by 11 ten-point rating scales. These rating scales were labelled with the 11 personality traits mentioned before, including the adjective (or adjectives) used in the description of the stimulus person. The rating scales ranged from 1 to 10, with end points denoted by "extremely not ..." and "extremely ..." (e.g. "extremely not *cooperative*" and "extremely *cooperative*"). In the two tasks subjects rated a total of 31 stimulus persons on 11 criterion variables.

The data were re-analyzed by Van der Kloot and Kroonenberg using three-mode principal component analysis. The solutions were based on double-centred subject matrices (see section 6.5). Thus the subjects were made identical with respect to scale and stimulus means, leaving the configurational aspects of the ITP's (i.e. the stimulus x scale interactions) as the data to be analysed.

10.3 THEORETICAL SUBJECTS *)

The data set was extended with six *theoretical subjects* in order to improve the interpretability of the solutions. The first, or *average subject* (A1), consists of the mean ratings of the stimuli averaged over the 59 real subjects. The second, or *dominance subject* (A2), has been constructed as if the stimuli were judged only with respect to their apparent dominance. The third, or *evaluation subject* (A3), was constructed as if the subject only judged the evaluative content of the stimuli. The fourth, or *random subject* (A4), consists of uniform random error superimposed on the overall scale means. The data of the fifth, or *uniform scorer* (A5), are equal to the grand mean, i.e. the average over stimuli, scales and subjects. The ratings of the sixth, or *extreme scorer* (A6), consists of either 2 or 9 scores. His scores are equal to 9 when the ratings of the average-subject (after being centred) are larger than 0. His scores are 2 when the average-subjects's double-centred ratings are smaller than 0.

---

*) Van der Kloot & Kroonenberg (1982) used the term *artificial subjects*, but *theoretical subjects* seems to be a better term.

We will use the theoretical subjects as some other authors have used 'conceptual individuals' or 'idealized individuals' (e.g. Tucker & Messick, 1963; Cliff, 1968; Tucker, 1972). The advantage of our theoretical subjects is that they were created on the basis of possible scoring behaviours of individuals from, or in analogy with, the original data. When included in the analysis they therefore provide a priori information about the subject space which is not the case with 'conceptual individuals'. The interpretation on the basis of theoretical subjects thus rests on more solid ground.

10.4 SCALE AND STIMULUS CONFIGURATIONS

The two-dimensional TUCKALS3 configuration of the *scales* explained 51.3% of the total sum of squares. Since the addition of a third dimension reduced the residual sum of squares by only 4.0%, we found the two-dimensional solution quite satisfactory, especially because the first two dimensions of the three-dimensional solution were virtually identical to those of the two-dimensional configuration. The configuration of the rating scales is pictured in Figure 10.1. It should be noted that each dimension explains an almost equal amount of variation: respectively 26.2% and 25.2%. This means that they are of equal importance for the group as a whole. The shape of the T3 configuration is roughly circular, and the horizontal and vertical dimensions can be interpreted as a dominance-submission and an evaluation dimension.

The T3 configuration of the *stimuli* for the 2x2x2-solution is represented in Figure 10.2. The 11 stimuli, consisting of single adjectives, lie on a polygon which is more or less the same as that of the scales in Figure 10.1, with the exception of *intelligent* and *industrious* which have switched places. Notwithstanding this difference, one may conclude that the stimulus space and the scale space are virtually identical; these spaces and their respective dimensions seem to have the same cognitive structure. Moreover, the two dimensions of the stimulus space also account for an almost equal proportion of the sum of squares (resp. 26.1% and 25.2%). Therefore

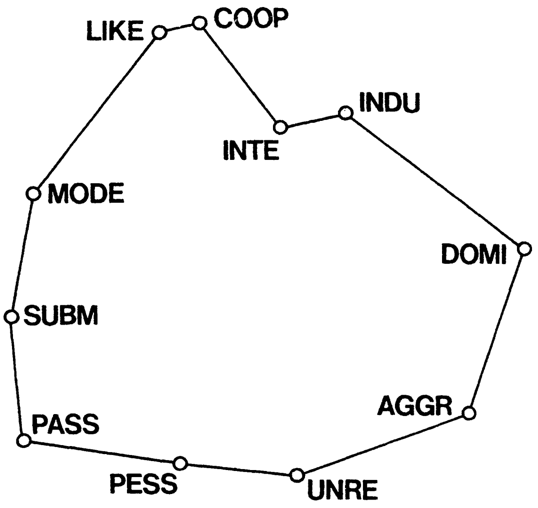


Fig. 10.1 *ITP study: Scale configuration*

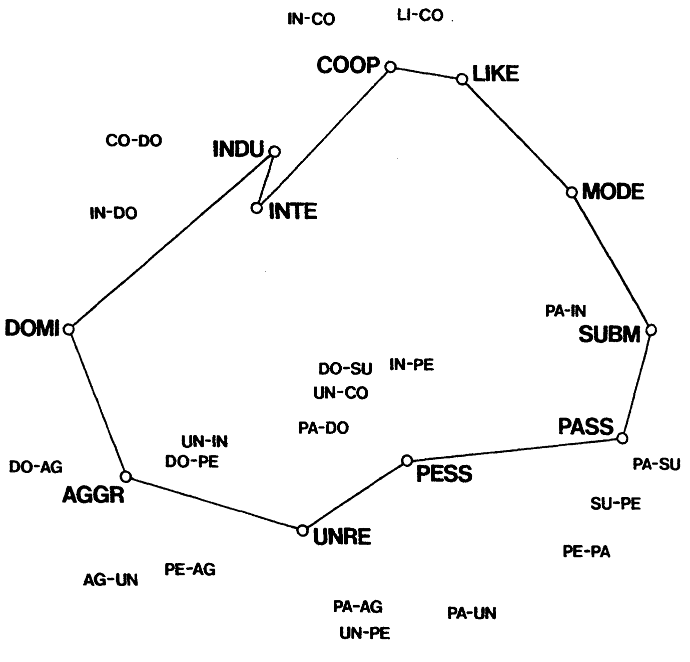


Fig. 10.2 *ITP study: Stimulus configuration*

Figure 10.1 and Figure 10.2 may directly be superimposed (after reflection) without further standardization of the projections, and we will refer to both the scale and the stimulus space as *the (personality) trait space.*

10.5 SUBJECT SPACES

The eigenvalues of the two components of the subject space from the T3 analysis (based on 65 real and theoretical subjects) were .498 and .015 respectively. Since the first component, which reflects the covariance of the individuals, is much larger than the second, it may be concluded that the subject space is largely one-dimensional (Fig. 10.3).

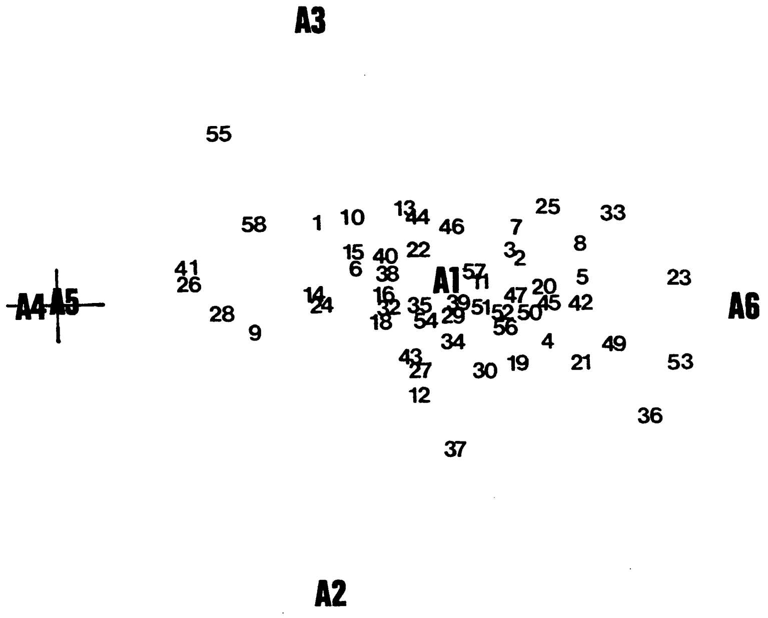


Fig. 10.3 *ITP study: Subject space* (+ = origin; unscaled)

The usefulness of introducing theoretical subjects now becomes clear, as they mark the end-points of the axes. The *uniform scorer* (A5) and the *extreme scorer* (A6) demarcate the first axis, and the *dominance subject* (A2) and the *evaluation subject* (A3) take on the extremes of the second axis. The *average subject* (A1) is located in the middle of the configuration. The *random subject* (A4) differs only marginally from the *uniform scorer* (A5). The conclusions (drawn on the basis of the theoretical subjects alone) are that subjects along the first axis of the subject space emphasize the dominance and evaluation axes of the personality trait space equally strongly with increasing emphasis going from left to right, and that subjects along the second axis of the subject space emphasize dominance at the cost of evaluation or vice versa.

The T3 core matrix tells the same tale (see Table 10.2). The diagonal elements of the core plane belonging to the first subject component have equal sizes and the same sign, and indicate therefore that both dominance and evaluation are weighted equally. The diagonal elements of the second subject component also have equal sizes, but opposite signs, indicating that either dominance or evaluation is emphasized. The larger size of the elements of the first core plane is a direct reflection of the larger eigenvalue of the first subject component.

Table 10.2 *ITP study: T3 frontal planes (two subject components and average frontal plane)*

71.89	- 2.18	- 9.20	- 2.14	31.34	- 2.16
2.19	70.39	- 1.78	9.39	.20	39.89
component 1		component 2		average plane	

Although the T3 core matrix supplies information how dominance and evaluation are weighted in relation to each other, it does not tell the size of such weightings for individual subjects. Such information is present in the loadings of the subjects on the two subject components. The extended core matrix from a TUCKALS2 (T2) solution supplies additional and more detailed information: the diagonal elements of each T2 core matrix indicate the amount of stretching and shrinking each subject applies to the axes of the common personality space, and the off-diagonal elements indicate the angle under which these axes are 'seen' (see section 6.9). It appears that all subjects see these axes as more or less orthogonal because the off-diagonal elements are never really large. Subjects with small and equal diagonal elements in their T2 frontal plane lie on the left hand side of the first axis of the subject space. Subjects with large and equal diagonal elements lie on the right hand side of the first axis, etc. Of the subjects who score most extremely (23, 37, 41 and 55) the core planes are shown in Table 10.3, along with an average subject (47), and the average core plane.

Table 10.3 *ITP study: T2 frontal planes (five subjects and average frontal plane)*

14.83	-.61
.03	13.34

subject 23

6.45	-2.79
-1.78	11.23

subject 37

3.30	.57
.04	2.14

subject 41

10.13	.36
-.58	9.99

subject 47

7.79	.13
-.36	-.57

subject 55

9.10	.02
.00	8.67

average plane

The most important feature of the T3 subject space is thus that most individuals emphasize the dominance and evaluation axes equally but with varying values of the weights. This implies that for most subjects the recovered personality trait configuration (or ITP) is circular, and that some have larger circles than others. The subjects with large weights (wider circles) have large sums of squares and thus use most of the ten-point scales. Of secondary importance is that some subjects emphasize either dominance or evaluation. Extreme examples are 55 and 37, who seem to use either the dominance or the evaluation axis as is confirmed by their T2 core plane in Table 10.3.

10.6 RESIDUAL/FIT RATIOS

Theoretical subjects can also assist in checking the assumption that all subjects only applied the transformations allowed by the model to the personality trait space. Subjects who do not conform to the model, or at least less so than other subjects, should have a smaller residual/fit ratio than the average individual. In the sums-of-squares plot for subjects (Figure 10.4) the contributions to the fit are plotted against the residual sums of squares for all real and theoretical subjects. The heavy line in this figure connects points with the overall residual/fit ratio (.49/.51). The other two lines connect points with ratios .39/.61 and .59/.41 respectively. These lines serve as a kind of confidence bands for the overall residual/fit ratio (see section 7.6).

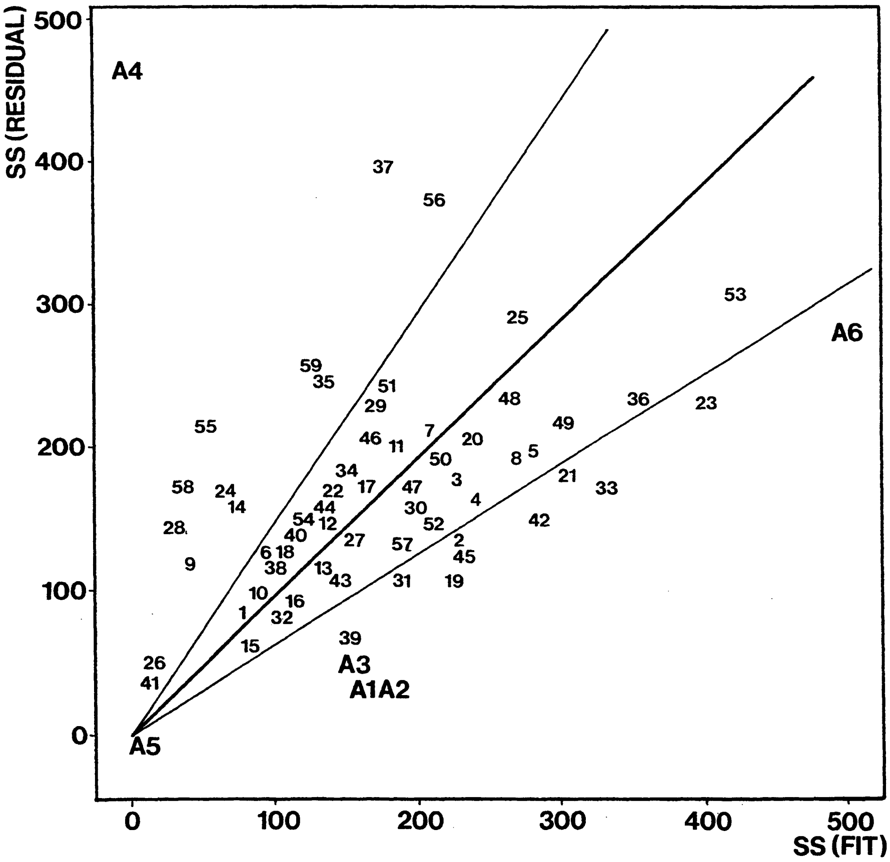


Fig. 10.4 *ITP study: Sums-of-squares plot for subjects*

Twelve real subjects (9, 14, 24, 26, 28, 35, 37, 41, 55, 56, 58, and 59) have rather large residual/fit ratios, and probably do not meet the assumptions of the model. Inspection of their scores on the subject components showed that: (a) the two subjects who have the most extreme scores on the second component space are ill-fitting points; (b) the majority of the 'bad' points have small values for the diagonal elements of their T2 core planes (for some examples see Table 10.3); (c) there are points which do not conform to this pattern, notable 35, 56, and 59, the 'best' of 'bad' points.

The 'good' points (2, 19, 21, 23, 31, 33, 36, 39, 42, and 45) generally have large scores on the first component of the subject space, and thus large overall sums of squares. It is, of course, no surprise that subjects with large sums of squares fit better than subjects with small sums of squares.

In Figure 10.3 the theoretical subjects also lie on the boundary of the configuration, and can thus be used to evaluate the real subjects. The *random subject* (A4) has practically no fit, as it should be. The *average subject* (A1) has roughly the same fit as a real subject (i.e. 17 in the centre of the subject space), but due to the averaging procedure, a smaller residual than such an individual. The *dominance subject* (A2) and the *evaluation subject* (A3) were created from the *average subject* (A1) with comparable sums of squares, which explains their position in Figure 10.3. The *uniform scorer* (A5) has fit nor error as his sum of squares is necessarily zero. The *extreme scorer* (A6) has understandably a very large sum of squares, and also a smaller residual/fit ratio (.36/.64) than the overall one (.49/.51), which indicates his scoring pattern is admissable in terms of the model. In fact, his private trait space is almost a perfect circle.

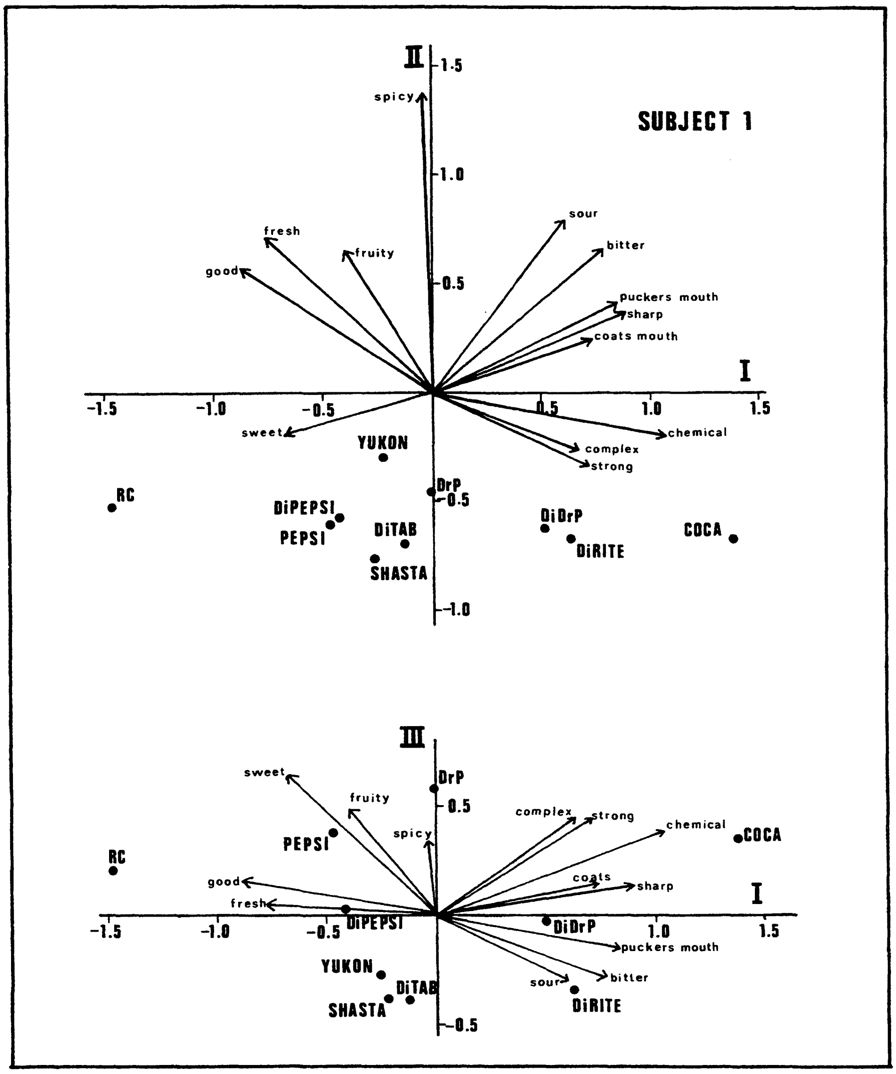
10.7 CONCLUSIONS

In conclusion one may say that it is especially useful to specify 'ideal' response patterns of theoretical subjects on the basis of a substantive theory, instead of the results of the analysis. By assessing the difference between the real and 'idealized' subjects, it is possible to accept or reject the models underlying the construction of the latter, and simplify the interpretations of axes.

SIMILARITIES AND ADJECTIVE RATINGS

11

cola study



11.1 INTRODUCTION

Schiffman, Reynolds, & Young (1981) discuss and apply a wide variety of multidimensional scaling programs and methods (e.g. INDSCAL, Carroll & Chang, 1970; SINDSCAL, Pruzansky, 1979; ALSCAL, Takane, Young, & De Leeuw, 1977; to a variety of data sets from sensory perception of smells and tastes. Two of these data sets, given in full by Schiffman et al. (p.33-36), refer to 10 different colas which have been compared with another *(similarity set)*, and which have been rated on 13 adjectives *(adjective set)* by 10 subjects. Thus the similarity set has dimensions: 10×10×10 (colas × colas × subjects) and the adjective set has dimensions: 10×13×10 (colas × adjectives × subjects).

Schiffman et al. derive stimulus spaces for the colas under various assumptions using a number of programs, and in their Chapter 11 they compare the solutions found. In a later section, 12.4, they use the adjective set to establish a relationship between the cola space from the similarity set and some of the adjectives. The latter is done in a rather roundabout way. In section 11.4 we will provide a proper analysis of the adjective set, and discuss ways to compare the results from the two sets. But first we will compare TUCKALS2 results on the similarity set with the results from other programs, and make a number of comments with respect to analysing (dis)similarity data via three-mode principal component analysis.

In contrast with most of our other examples we will not show many results graphically, since Schiffman et al. have already done so, especially for the similarity set. Furthermore, we will not go into the experimental procedures followed to collect the data; they are described in Schiffman et al.'s section 3.1. Finally, we will frequently use the terminology connected with multidimensional individual differences scaling and optimal scaling without explaining their meaning in detail. For this we again refer to Schiffman et al.(p.14-17).

11.2 SIMILARITY SET

Data. The similarity set is presented by Schiffman et al. (p.34) as 10 lower triangular matrices of dissimilarities, which we have converted to similarities; the diagonal elements of each matrix were given a similarity of 100. In three-mode principal component analysis it is preferable to analyse similarities rather than dissimilarities. If we assume that we have symmetric input data, in a Tucker2 analysis G and H will necessarily be the same. However, if high values indicate dissimilarities, the first components of G and H are inversely related, because highly positively loading stimuli are very 'not dissimilar'. This means that the points indicating the same stimulus have diametrical positions when constructing a joint plot for the first and second modes. This is, of course, rather confusing for interpretation. This becomes even more so when the data are slightly asymmetric, because in that case one is explicitly interested in the closeness of the stimulus as row and as column variables. Therefore, in other kinds of data, too, high values should indicate that the row and column object have much in common.

After the conversion to similarities the data were double-centred (jk,ik-centring, see section 6.5) to make them resemble the scalar-product input used, for instance, in INDSCAL. One could argue from the point of view of classical Torgerson (1958) multidimensional scaling that first squaring the entries would be more appropriate, but as the loss function is already quadratic, squaring the entries would emphasize the larger distances overly much.

Cola spaces. In line with the Schiffman et al.'s analyses a Tucker2 solution with 3 components each for rows and columns, was determined. It was attempted to improve diagonality of the core matrix by orthonormal and non-singular transformations (see Chapter 5), but the improvements in diagonality were rather small, mainly due to the near-diagonality of the original core matrix.

Table 11.1 *Cola study (sim): comparisons between cola spaces*

	1	2	3	4	5	6	7	8	9	10
component 1 (diet - nondiet)*										
(S)INDSCAL	-.35	.29	.21	.21	.31	.21	-.32	-.39	.31	-.47
ALSCAL ƒ	-.34	.29	.12	.21	.38	.18	-.24	-.44	.33	-.46
TUCKALS2 $	-.35	.28	.22	.19	.31	.24	-.33	-.38	.30	-.47
TUCKALS2 £	-.35	.27	.21	.21	.31	.23	-.31	-.39	.30	-.47
TUCKALS2	-.31	.32	.18	.08	.33	.24	-.42	-.36	.35	-.42
component 2 (cherry-regular)*										
(S)INDSCAL	.18	.24	-.10	-.63	.16	.12	-.57	.14	.27	.20
ALSCAL ƒ	.25	.28	-.33	-.53	.18	.08	-.53	.08	.31	.21
TUCKALS2 $	.20	.26	-.17	-.60	.18	-.05	-.56	.09	.31	.23
TUCKALS2 £	.20	.24	-.15	-.62	.17	.07	-.55	.12	.29	.24
TUCKALS2	.21	.13	-.10	-.69	.10	.09	-.48	.25	.18	.32
component 3 (manufacturer's flavour)*										
(S)INDSCAL	.36	.32	-.55	.29	-.04	-.36	.10	-.41	.26	.03
ALSCAL ƒ	.35	.26	-.48	.35	-.13	-.47	.25	-.35	.16	.07
TUCKALS2 $	.24	.28	-.51	.42	.02	-.44	.14	-.40	.23	.01
TUCKALS2 £	.29	.31	-.55	.32	.03	-.45	.07	-.36	.26	.08
TUCKALS2	.33	.34	-.57	.22	.05	-.44	-.01	-.33	.30	.13

ƒ interval-continuous and scaled to unit length (i.e. divided by $1/\sqrt{10}$), component 2 and 3 interchanged.

$ after non-singular transformation of core matrix: TUCKALS2-NS.

£ after orthonormal transformation of the core matrix: TUCKALS2-ON

* for an explanation of the labelling of the axes see Schiffman et al., p. 218. The colas are ordered as in Schiffman et al. (Table 3.2).

In Table 11.1 comparisons are given between the transformed and untransformed TUCKALS2-solutions and those from three individual differences techniques. The conclusion from this table is straightforward: all solutions are virtually identical with the largest differences in the last component. That the much less restricted T2 model hardly improves the fit to the data, is the result of the very nice agreement of the data to the more restricted models. This follows, for instance, from a mere loss of 4% variation accounted for when the off-diagonal elements from the T2 core planes are set to zero.

Importance of dimensions. Difficulties arise when we try to assess the relative importance of the dimensions. (S)INDSCAL uses an approximate percentage of variance accounted for, ALSCAL offers the average subject weight per dimension, and TUCKALS provides a variation accounted for, and an average subject weight. Schiffman et al.'s comparison (p.243) of ALSCAL average subject weights with (S)INDSCAL variation (or variance) accounted for is inappropriate, as it amounts to comparing something like a variance with someting like a standard deviation. In TUCKALS the component weights (after double-centring) are comparable with (S)INDSCAL and ALSCAL approximate variance accounted for, and the TUCKALS average subject weights, i.e. the diagonal elements of the average T2 core plane, are comparable to the ALSCAL average subject weights (provided the former are divided by n, the number of subjects). After all, the total rescaled variation present in the TUCKALS2 analysis is 1000 (ℓ×m×n = 10×10×10), while in ALSCAL (and in (S)INDSCAL) this variation is 10 (nxl = 10xl) due to the matrix-conditional equalization of subjects' variances to 1 (k-standardization; see section 6.6). In Table 11.2 the various comparisons are given. Note that the ALSCAL values are now in far better agreement with (S)INDSCAL than in Schiffman et al.

Subject weights. The comparison of the subject weights or saliences (Table 11.3) is relatively straightforward, and most of the analyses agree. The largest discrepancies are to be found in the ALSCAL analysis in which the subject weights lie generally below those found with TUCKALS and (S)INDSCAL.

Table 11.2 *Cola-study (sim.): importance of components*

	"approximate variation accounted for"			average subject weight		
(S)INDSCAL	.29	.23	.11	.48	.42	.32
ALSCAL *f*	.22	.11	.11	.44	.33	.31
TUCKALS2	.32	.22	.11	.49	.39	.31
TUCKALS2-ON	.31	.21	.11	.49	.39	.31
TUCKALS2-NS	.31	.21	.11	.49	.38	.32

f interval-continuous

. ALSCAL (appr. variance accounted for) = $\Sigma(\text{subj.weight})^2/10$
. TUCKALS2 average subject weights have been divided by 10 (see text)

The TUCKALS2 weights are the diagonal entries of the TUCKALS2 core planes, in other words the off-diagonal entries are disregarded, but this is of little importance as they were small anyway. The loadings of the TUCKALS3 subject space are included in Table 11.3 to allow comparison with the TUCKALS2 results. It is clear that the TUCKALS3 loadings more or less follow the sizes of the diagonal elements, but the agreement is not perfect. We will not attempt a detailed analysis of these here.

According to Schiffman et al. the subject weights for the first and second stimulus components (i.e. diet vs. non-diet, and cherry vs. regular, respectively) discriminate almost perfectly between PTC-tasters and non-tasters (p.151, and p.306 respectively) "PTC is a compound which tastes bitter to some and is tasteless to others" (p.151), and "The ability to taste PTC is related to one's perception of artificially sweetened drinks (Four of the colas [i.e. the Diet ones] are artificially sweetened)" (p.305).

With respect to the third component, it is not clear why subjects 2,4,7 and 9 are especially sensitive to 'manufacturer's flavour'.

Fit of subjects. Comparing the fit of a subject from different programs is not an easy task, mainly because each program has its own definition of fit, or at least its own terminology.

Table 11.3 *Cola study (sim.): subject weights*

	component 1					(diet - nondiet)*				
	PTC tasters					non-PTC tasters				
	5	6	4	1	9	3	7	2	10	8
(S)INDSCAL	.81	.80	.66	.64	.62	.20	.21	.19	.21	.43
ALSCAL f	.62	.60	.57	.44	.47	.29	.38	.33	.32	.43
TUCKALS2-NS	.88	.82	.69	.67	.64	.21	.22	.17	.20	.37
TUCKALS2-ON	.88	.82	.69	.66	.65	.21	.22	.17	.20	.37
TUCKALS2	.83	.80	.65	.70	.65	.22	.28	.16	.24	.39
TUCKALS3 -1	.34	.33	.36	.35	.33	.28	.30	.28	.26	.30
	component 2 (cherry-regular)*									
(S)INDSCAL	.07	.09	.25	.35	.20	.74	.65	.58	.69	.56
ALSCAL f	.18	.20	.24	.26	.22	.54	.36	.35	.37	.38
TUCKALS2-NS	.25	.27	.44	.26	.41	.26	.38	.35	.26	.34
TUCKALS2-ON	.03	.06	.22	.32	.19	.68	.67	.65	.62	.45
TUCKALS2	.10	.10	.28	.31	.19	.65	.61	.62	.61	.46
TUCKALS3 -2	-.44	-.39	-.19	-.11	-.19	.39	.36	.39	.33	.13
	component 3 (manufacturer's flavour)*									
(S)INDSCAL	.24	.27	.43	.24	.40	.29	.33	.39	.22	.40
ALSCAL f	.30	.31	.37	.28	.33	.38	.33	.33	.30	.37
TUCKALS2-NS	.25	.27	.44	.26	.41	.26	.38	.35	.26	.34
TUCKALS2-ON	.23	.26	.42	.24	.41	.27	.37	.39	.23	.31
TUCKALS2	.20	.24	.40	.21	.40	.29	.37	.43	.21	.29

f interval-continuous

(S)INDSCAL computes correlations between computed scores and the scalar products derived from the individual subjects' data; these are equal to the square root of the sum of squared subject weights, although the program does not say so. As such they can be seen as a measure of fit, because the total sum of squares of the scalar products of each subject has been scaled to one. Schiffman et al. are very vague as to the exact nature of these correlations.

ALSCAL likewise computes (squared) correlations which are equal to the sum of the squared subject weights for matrix-conditional data, which is the only case treated here. Note that also in

ALSCAL the total sum of squares of the scalar products for each subject has been scaled to one. Thus the ALSCAL and (S)INDSCAL correlations should be comparable.

However, Schiffman et al. (p.175) state with respect to ALSCAL that "the squared correlations RSQ show the proportion of variance of the disparaties accounted for by the MDS model", which sounds like something different from the definition given above for (S)INDSCAL. They continue to say that "ALSCAL correlations are not calculated in the same way as INDSCAL correlations. This point is discussed in Chapter 16", but a comprehensive treatment of this point cannot be found in that chapter.

TUCKALS has as its measure of fit for subjects the fitted sum of squares per subject, which can be normed by (1) the average SS(Tot) per subject giving us the ***Proportional SS(Fit)***, or (2) a subject's total sum of squares, giving us the ***Relative SS(Fit)*** (see Chapter 7). The two are different because in TUCKALS no equalization of variation per subject (or elements of the third mode) is performed internally.

Table 11.4 *Cola study (sim.): fit of subjects*

	6	5	4	3	8	1	7	9	10	2	over-all
(S)INDSCAL ss	.71	.71	.69	.67	.67	.59	.58	.59	.57	.53	.63
ALSCAL ƒ,ss	.49	.51	.52	.52	.46	.34	.38	.38	.33	.34	.43
TUCKALS2 p	.74	.74	.70	.66	.61	.62	.61	.62	.60	.57	.65
TUCKALS2 r	.75	.84	.72	.60	.46	.66	.67	.65	.51	.64	
TUCKALS2 d	.70	.74	.66	.56	.45	.62	.59	.62	.47	.60	
ALSCAL s,f	.78	.78	.80	.78	.79	.76	.77	.77	.75	.76	.77

ƒ= interval-continuous
ss =sum of squared subject weights
p = proportional SS(Fit)
s = 1 - stress
r = relative SS(Fit)
d = relative SS(Fit) from diagonal elements only

Table 11.4 shows that in general both the Prop. SS(Fit), and the Rel. SS(Fit) are higher than the (S)INDSCAL SS(subj.weight)s, as one would expect with a less restricted model. Due to the lower subject weights in ALSCAL, the fit of the subjects is also lower than those in the other analysis. In Table 11.4 we have included '1-stress' for each subject but this is unlikely to be a correct measure for comparison, because it has a non-linear relation to the squared correlation RSQ.

11.3 ADJECTIVE SET

Data. As mentioned in section 11.1 the 10 subjects also rated the colas on 13 continuous scales by checking a 5-inch line of which the end points were marked. The positively marked adjectives were placed at the left hand side, scored as 0, and their negation at the right hand side, scored as 100 (see Schiffman et al., p.31). For our analysis the scores were inverted. In this way the scoring remained in line with the similarity set, where the dissimilarities were converted in similarities. Furthermore, the data were centred by subtracting the midpoint of the scale (50) from the inverted scores. By this centring overall differences between judgements of subjects with respect to, for example, the sweetness of colas were maintained in the analysis, as well as any differences between the colas themselves.

Schiffman et al.'s analysis. Schiffman et al. did not present an analysis as such of the adjective set, but they reduced the set directly by averaging over subjects, and by computing correlations between the adjectives over colas. On the basis of these correlations they chose a subset of 6 adjectives for their further analyses. Via a canonical regression procedure (using BMDP6M, cf. Dixon, 1981) they derived three canonical variates for the loadings on the cola dimensions from the similarity set jointly with the scores of the colas on the six selected adjectives.

The procedure contains clearly a number of ad-hoc decisions, e.g. exclusion of certain adjectives, and averaging over subjects. It bypasses the structure in the adjective set itself, and possible individual differences with respect to the usage of adjectives. The latter is somewhat strange as one of Schiffman et al.'s main conclusions from the comparisons of various techniques for dealing with three-way similarity data is that individual differences analyses are recommended whenever possible (p.251). For the adjective set the three-mode principal component model can be used to do justice to individual differences, as we will demonstrate shortly.

Tucker2 analysis. To stay in line with Schiffman et al. we decided to report a *3x3-solution* for the Tucker2 model, i.e. 3 components each for the colas and the adjectives. Although for presention some reordering of the colas would be preferable, we chose to maintain the original order in Table 11.5 to facilitate comparisons with the Schiffman et al.'s analyses.

Table 11.5 *Cola study (sim.): TUCKALS2 Cola space*

Cola	1	2	3	Diet	Type	Manu-facturer
Diet Pepsi	.08	.30	.17	+	regular	PC
Royal Crown Cola	-.17	.42	.56	-	regular	RC
Yukon	.36	-.01	.19	-	regular	YU
Dr. Pepper	-.34	.46	-.16	-	cherry	Dr.P
Shasta	.39	.25	.18	-	regular	SH
Coca Cola	.04	.38	-.68	-	regular	CC
Diet Dr. Pepper	.22	.24	-.21	+	cherry	Dr.P
Tab	.48	.16	.15	+	regular	CC
Pepsi Cola	-.19	.47	.09	-	regular	PC
Diet Rite	.51	.12	-.18	+	regular	RC
% explained variation	.25	.17	.11			

The most surprising fact about the cola space is that it hardly resembles the cola space from the similarity set. The major axes of the latter set were labelled by Schiffman et al. as 'diet/non-diet', 'cherry/regular', and 'manufacturer's flavour', but these axes are not to be found here. Closest to the earlier axes seems to be the first one here, which contrasts regular colas with diet colas plus the local ones (Yukon, Shasta). We will compare the solutions in more detail in the next subsection.

Table 11.6 *Cola study (adj.): Adjective space*

Adjectives[f]	components			components[$]		
	1	2	3	1	2	3
good (bad)	-.27	.34	.05			
strong (weak)	.18	-.15	.41			
sweet	-.32	.03	.47			
bitter	.38	.19	-.22			
sour	.33	.27	-.24	.53	.18	-.50
fruity	-.12	.37	.32	-.08	.53	.42
spicy	.10	.64	.18	.25	.67	.28
coats mouth	.28	.07	.15			
sharp	.35	.11	.14			
puckers mouth	.36	.09	-.08	.58	.02	-.22
fresh	-.22	.38	-.06	-.28	.44	-.47
chemical	.33	-.13	.38	.48	-.22	.48
complex (simple)	.17	-.12	.41			
% explained variation	.35	.13	.04	.30	.23	.04

[f] between brackets other marker; otherwise: not ......

[$] results from TUCKALS2 analysis with Schiffman et al.'s selection of adjectives; note that these loadings are larger because all components have unit lengths.

The adjective space shows a first axis which is more or less evaluative (good, fresh, sweet versus bitter, sharp, puckers mouth, chemical). The second component is marked by spicy, fruity, good, and fresh, and has only three slightly negative loadings. The third component is actually too small to consider it a valid descriptor for all the colas together and/or all subjects together. Table 11.6 also shows the results of a separate analysis using only the six adjectives selected by Schiffman et al. The loadings and the associated cola space (not shown) support the contention that the subset of six adjectives is representative of the entire set of adjectives.

For a proper insight into the nature of the adjective set it is necessary to investigate how the cola and adjective spaces are related. One way to do this would be to plot them jointly on the basis of the average TUCKALS2 core plane (internal averaging; see section 6.10), or one could average over subjects externally as

Schiffman et al. did. Both averaging procedures presuppose that all subjects use the adjectives to describe the colas in more or less the same way. In this data set this is definitely *not* the case. In fact, the differences between the subjects are probably larger than in any other example in the present book (see Table 11.7). To grasp the extent of these differences between the subjects in the adjective set, it is worthwhile to look back at Table 11.3.

Table 11.7 *Cola study (adj.): subjects*

£ Subject	elements of T2 core planes * (x10)									T3 subjects components (x100)			T2 Rel. SS(Fit) (x100)
	1,1	2,2	3,1	2,1	1,2	2,3	3,3	3,2	1,3	1	2	3	
9	65	-12	-43	- 2	-12	22	0	- 5	5	33	-17	-46	55
8	45	-23	-33	6	- 7	12	-12	-18	-11	27	-18	-26	46
5	91	21	26	-11	-15	1	0	- 2	4	42	46	-12	64
2	64	5	36	-34	7	3	- 4	-20	- 4	30	42	- 9	48
7	35	-36	-43	-49	- 7	7	-34	4	- 2	36	-39	13	40
1	10	-32	-55	-18	-24	9	-12	13	- 6	21	-55	5	31
6	41	-61	4	- 1	-22	- 4	- 6	9	-13	33	- 3	21	47
4	66	-33	13	-31	-26	-15	- 2	-16	-21	42	25	19	59
3	7	-58	12	-44	-40	12	-10	-11	2	29	-12	58	51
10	37	- 1	-21	19	1	38	- 5	-11	7	15	-13	-51	38
%explained variation	21	9	8	5	3	2	2	2	1	30	10	7	

\$ i,j = c_{ijk} = relation for kth subject.
* the core elements are arranged in decreasing order of importance
£ the subjects were grouped as far as possible with respect to their pattern in the core matrix.

For the similarity set (Table 11.3) the off-diagonal elements were not given as they were too small; the diagonal elements included in the table vary, but not dramatically so. For the adjective set these *same* subjects are practically all different. Surprising is also that now the distinction between PTC and non-PTC tasters is nowhere to be found, while this difference dominated in the similarity set. As the average core plane is not representative of any subject, we present here joint plots for subject *1* (Fig. 11.1A) with a relatively bad fit of .30, and subject *4* (Fig. 11.1B)

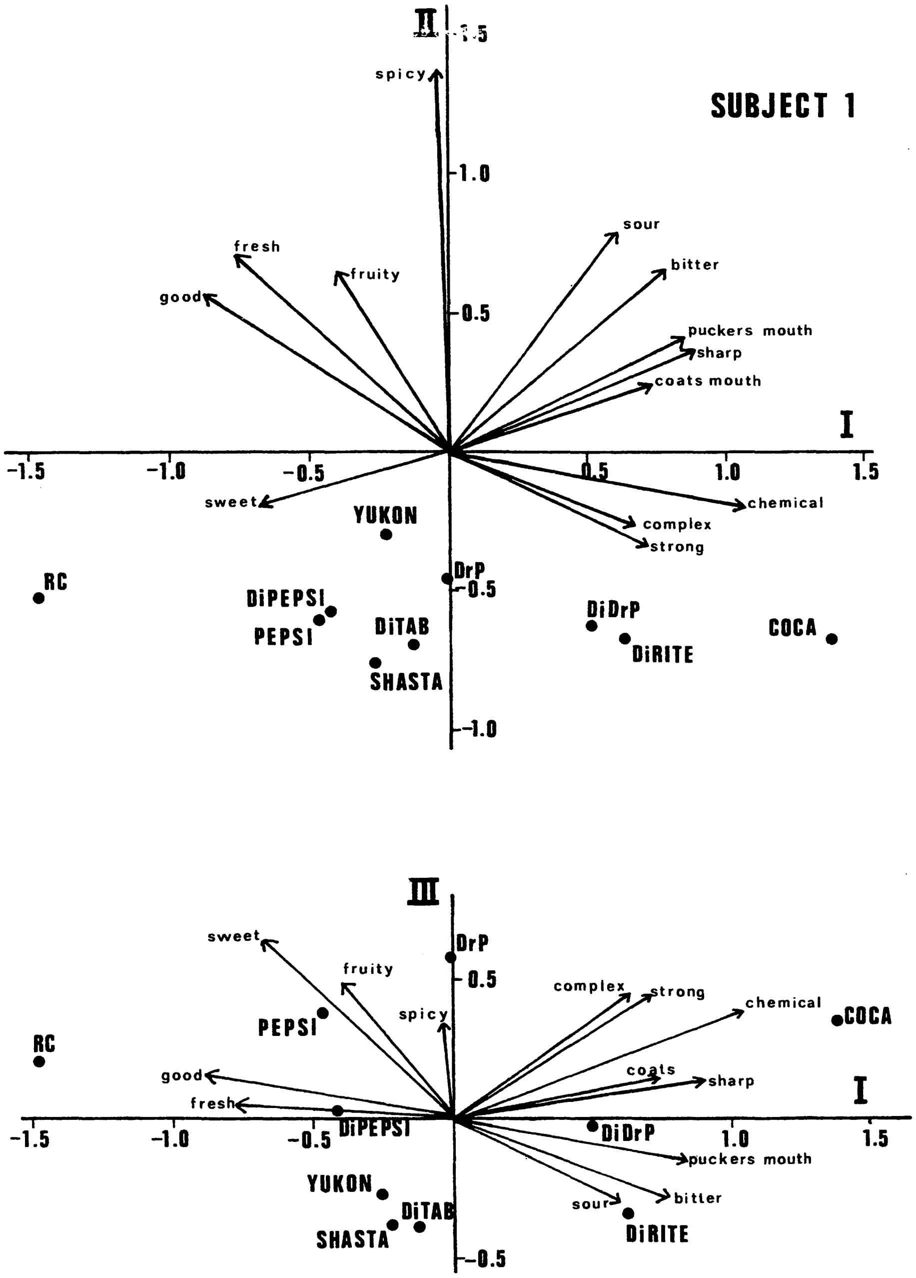


Fig. 11.1A *Cola study (adj.): Joint plots for subject 1*

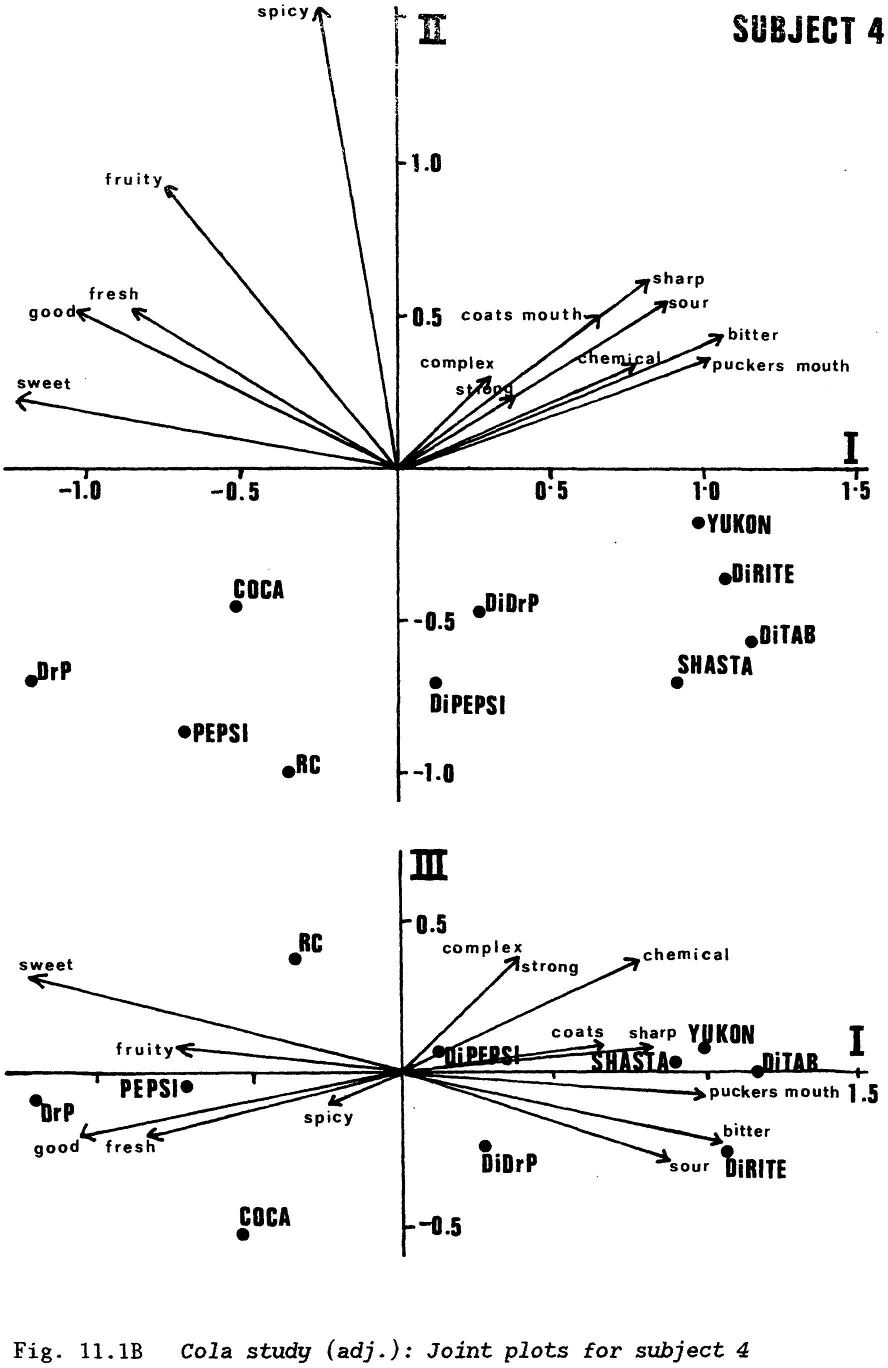


Fig. 11.1B *Cola study (adj.): Joint plots for subject 4*

with a relatively good fit of .60.

By looking at the projections of the colas on the adjective vectors, or vice versa, we see that:

1. To subject *1* only Dr. Pepper, Pepsi Cola, and Coca Cola, and to subject *4* only RC Cola taste moderately good. This is hardly surprising, for who likes colas which have been standing around for two hours (Schiffman et al., p.31). Similarly one cannot expect them to taste very fresh.
2. Subject *1* and *4* disgree absolutely about the taste of Coca Cola. To *1* it tastes very chemical, strong, and complex, and definitely not good, fresh, fruity, or sweet. To *4* it tastes sweet, good, not sharp, sour, or bitter. Similar differences can be found with respect to other colas, and other subjects.
3. On the whole (looking at the first axes only), one might say that *1* likes RC Cola, Pepsi Cola, Diet Pepsi, Yukon, and Shasta, while *4* likes Dr Pepper, Coca Cola, Pepsi Cola, and RC Cola, and abhors Diet Colas and local brands. The appreciation might, of course, be different when the colas are not served tepid and decarbonated.

Overlooking the analysis as a whole, one wonders if the data do not contain too much error. Only half (.52) of the variation of the centred data could be explained in the TUCKALS2 analysis, and noting the vast differences in usage of the adjectives, one cannot help wondering if the adjectives meant the same thing to all subjects. The fact that no two subjects agreed in their use of the dimensions of the cola and adjective spaces, while they reacted consistently on the similarity rating task also points in the same direction.

11.4 SIMILARITY AND ADJECTIVE SETS

Schiffman et al. used the adjective set in a canonical correlation analysis to find out whether the structure in the cola space from the similarity set (or similarity cola space, for short) could be associated with specific adjectives. The difficulty with their approach to this problem is that the adjective set had to be moulded to fit the analysis.

Above we already remarked that their procedure is rather ad-hoc, and we also noted that there are such large individual differences in usage of the adjectives that one might doubt if a valid answer can be obtained with Schiffman et al.'s procedure for the subjects as a group.

Our first solution to the problem as described in the previous section, was simply to analyse the adjective set itself to find out how the subjects used the adjectives to describe the colas. This led to a space for the colas considerably different from the similarity cola space, and to a different evaluation of the subjects. One might argue that our approach is, however, not an answer to the question posed. After all, we did not use the similarity cola space at all.

In principle it is possible to tackle the problem directly within the TUCKALS framework by a so-called external analysis (see section 3.7, and Carroll, 1972, for a discussion of external analyses in multidimensional scaling). For such an analysis a three-mode principal component analysis would be performed on the adjective set with the components of the cola mode fixed at the loadings of the similarity cola space. By using the TUCKALS approach due regard is paid to the individual differences, in contrast with Schiffman et al. In a sense one can look upon such an analysis as an individual differences canonical regression analysis. Unfortunately the TUCKALS programs have not yet been adapted for such analyses, but this will be done in the future. Incidentally, similar options for external analysis are included in ALSCAL-4 (Young & Lewyckyj, 1979), and ALSCOMP3 (Sands & Young, 1980).

A second less direct approach would be to link the cola space from the adjective set (or adjective cola space) to the similarity cola space by a procrustes rotation (see e.g. Gower, 1975). The correlations between the similarity cola space and the rotated adjective cola space may then serve as measures of agreement. The results of such a procedure are given in Table 11.8. At the outset it was not clear to us whether an adjective cola space based on ratings centred at 50, or one based on double-centred ratings would

Table 11.8 *Cola study (sim. + adj.): correlations between cola spaces*

		rotated axes of adjective cola space (*ACS*)					
		centred at 50			double-centred		
		1	2	3	1	2	3
axes of	1	.50	.16	.49	.44	.18	.37
similarity	2	.23	.43	.07	.18	.62	-.09
cola space	3	.55	.05	.71	.37	-.09	.69
(*SCS*)							

Note: The correlation between the unrotated axes 1 and 2 of *ACS* is -.88; the correlation between the rotated axes 1 and 2 of *ACS* is .92 for the data centred at 50. In the double-centred set all such correlations are less than 1/10000.

be best to use for comparison; therefore, both are included. The conclusion from the correlations between the axes of the two spaces is that their structures are different, even after rotation, irrespective of the centring used.

In sharp contrast with this conclusion is the result of Schiffman et al. (p.297) which shows via canonical correlation analysis that their reduced set of adjectives and the similarity cola space are rather strongly related. It turns out that this result is

Table 11.9 *Cola study (sim. + adj.): canonical correlation analysis*

		loadings					
		centred at 50			double-centred		
		1	2	3	1	2	3
axes of	1	.56	-.25	-.79	.60	-.23	-.77
similarity	2	.11	-.92	.37	.10	-.93	.36
cola space	3	.82	.29	.49	.79	.30	.53
can.corr.		.88	.69	.08	.95	.69	.10
redundancy		.26	.16	.00	.30	.16	.00
significance		.22	.47	.85	.04	.47	.81

partly a method effect. This can best be illustrated by applying a canonical correlation analysis (BMDP6M; Dixon, 1981) to our two cola spaces (Table 11.9). We here present only the canonical variate loadings for the similarity cola space, as this space is the target. From Table 11.9 we certainly observe more 'promising' results than from the procrustes rotations, although still only in the double-centred case one of the axes reaches significance. Note, however, that Schiffman et al. (p.191), advise using procrustes rotations for comparisons of stimulus (or distance) spaces, rather than canonical correlation analysis.

The better fit of Schiffman et al. thus seems to result from the use of canonical correlation analysis, and from their more direct approach to their own question whether the adjectives can be associated with the axes of the similarity cola space, provided it is correct to use the averaged ratings. The problem is that they bypass the structure of the adjective set itself. Thus, although on the average adjectives can be associated with the axes of the similarity cola space, the wide differences between the two cola spaces and between the importance which subjects attach to the axes, stand out as a major result of the comparison.

11.5 CONCLUSION

In conclusion it seems fair to say that although the ten subjects perceived the colas in the same way when judging their similarities (except for the easily explained PTC-nonPTC difference), they were unable to give descriptions of the colas that were consistent with their similarity judgements.

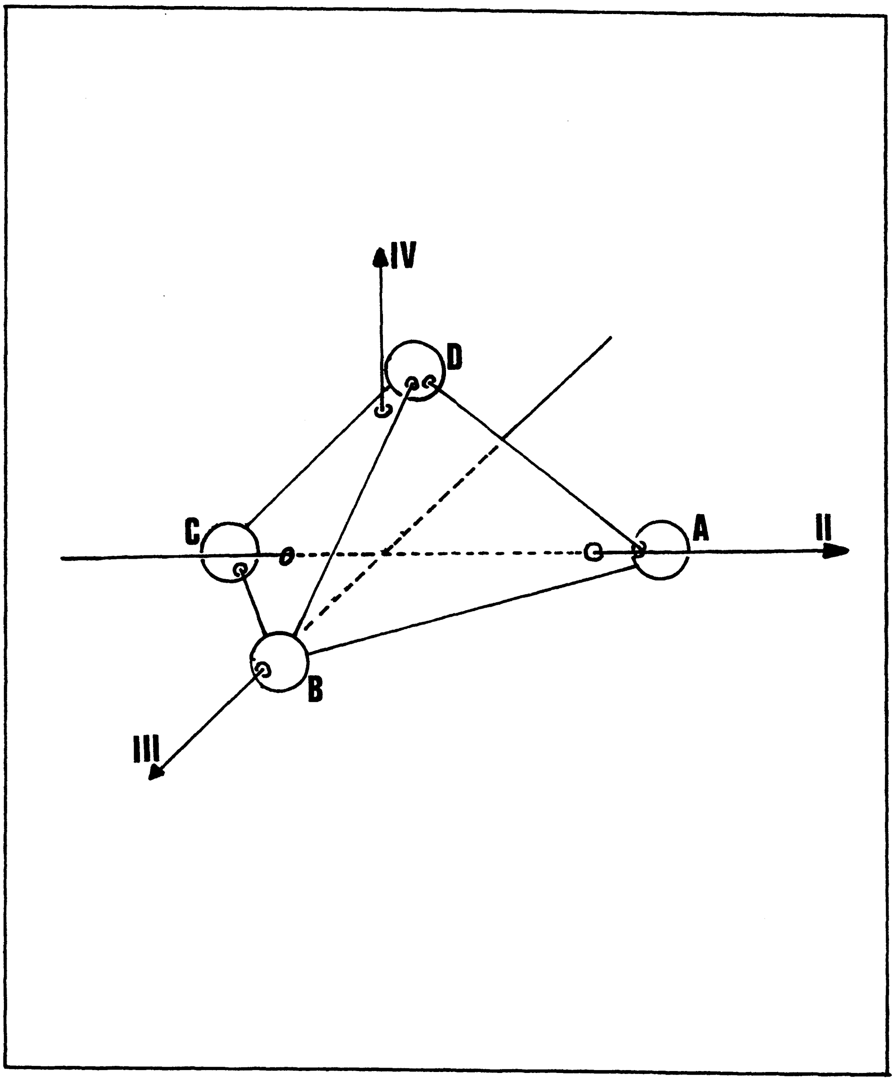
This conclusion almost echoes a conclusion reached in another experiment on sensory perception, namely on distortion of sound. In the words of Kruskal (1978):

> The most significant aspect of her [Mc Dermott's, 1969] article was the unequivocal evidence it provided that listeners differ strongly on the evaluation of distortion, even though (as indicated by the multidimensional scaling) they perceive the distortion in much the same way. This raised serious questions about the previously conventional method of averaging preferences across listeners, and ignoring individual differences. (p.320).

CORRELATION MATRICES

12

four ability-factor study



12.1 INTRODUCTION

In 1976, Glass issued a plea for a moratorium on data collection in favour of performing 'meta-analyses' on published research already available. His argument was that much information was available on many subjects, but that very few attempts had been made to integrate the available results. Review papers generally aim at citing research, rather than at re-analysing and integrating findings. In this chapter we will show how three-mode principal component analysis can be used for 'meta-analysis' on correlation matrices taking data from Meyers, Dingman, Orpet, Sitkei, & Watts (1964) as an illustration.

Especially in the field of intelligence tests, correlation matrices are computed for the groups used for calibrating the tests. For test developers it should be important to know whether the relationships between the subtests are the same for each calibration group. However, Wechsler (1974) for instance, published in his test manual eleven correlation matrices of subtests of the revised Wechsler Intelligence Scale for Children (WISC-R), one for each age group, without investigating the correlation matrices in a systematic fashion. In a later paper we hope to turn to an analysis of these and other correlation matrices of the WISC-R.

12.2 THREE-MODE ANALYSIS OF CORRELATION MATRICES

Using correlation matrices as direct input for a three-mode analysis introduces some complications compared to regular raw data. By taking correlation matrices, we treat the correlations as we do any raw input data, except that the input scaling poses no problems: no scaling is desirable or necessary. In fact, the data have already been scaled (jk-normalized; see section 6.5). One could treat all correlations matrices as equally important, or according to the number N of individuals in the sample. In the latter case one should multiply the correlations with $\sqrt{N}$.

Particular to the analysis of (published) correlation matrices is that one generally wants to compare the three-mode analysis with separate (two-mode) ones to assess their similarities and differences. The comparisons cannot be made directly as in three-mode analysis of correlation matrices the eigenvalues, or weights for the components, are quadratic functions of the eigenvalues from a standard principal component analysis. In other words, it is the square root of the standardized three-mode weights which should be compared with the proportion of explained variance of a regular principal component analysis.

We will only employ the Tucker2 model and not reduce the third mode, as we wish to compare the results from a three-mode solution with those of separate principal component analyses. Comparisons to assess how well the three-mode solution agrees with the separate analysis for each group can be made using the T2 core plane of that group. How this works follows from the observation that in a separate principal component analysis the correlation matrix of the k-th subgroup can approximately be decomposed as

$$R_k \cong H_k \Phi_k H_k' \qquad \text{with } \Phi_k \text{ diagonal, and } \phi_j^k \text{ the j-th eigenvalue, } j = 1,\dots,q \text{ ,}$$

and in a three-mode analysis

$$\tilde{R}_k \cong H\tilde{C}_k H' \qquad \text{with } \tilde{C}_k \text{ the k-th frontal plane, which is not necessarily diagonal.}$$

Given that H and H_k are similar enough to allow comparisons, the importance of the respective axes in the two spaces can be assessed using Φ_k and $\tilde{C}_k$. It should be kept in mind that the absolute size of the entries in Φ_k and $\tilde{C}_k$ depend on the lengths of the vectors in H and H_k. In TUCKALS H is columnwise orthonormal, and in many principal component analyses H_k has columns of lengths

ϕ_i^k. In other words, the values for the loadings in a standard principal component analysis correspond to a decomposition

$$R_k = \tilde{H}_k\tilde{H}_k' \text{ with } \tilde{H}_k = H_k\,\Phi_k^{\frac{1}{2}}\ .$$

Furthermore, in the TUCKALS programs the input data are generally rescaled such that the total sum of squares is $\ell \times m \times n$. In that case some adjustment is necessary before the actual comparison can be carried out.

12.3 OTHER APPROACHES

A number of other ways exist to deal with sets of correlation matrices. Within the general framework of analysis of covariance structures, Jöreskog (1971) has developed a method which he calls *simultaneous factor analysis for several populations*. Given a theoretical model for the factor loadings the correlation matrices may be analysed jointly to see if they all fit the same hypothesized structure. For the Meyers et al. data this seems an attractive alternative to the one presented here, as a clearly defined a priori structure is available. In studies in which this is not the case, the approach seems less easy to apply.

A second way to deal with this kind of data is to perform separate component analyses for each of the correlation matrices, and compare the component loadings or weights via the *perfect congruence approach* (Ten Berge, 1977; 1982). This approach assumes, as the previous one, that a target structure is available, either from a previous study or from theoretical considerations.

Finally, the approach taken by Meyers et al. in the analysis of their data may be employed, i.e. using the same *transformational procedure* on the component solutions, and hope that they point in the same directions. Alternatively, all components may be transformed via a procrustes rotation (see, e.g. Gower, 1975) to a common target. The former approach gives little guarantee that the desired result will be obtained, the latter is shown to be suboptimal by Ten Berge (1982).

12.4 FOUR ABILITY-FACTOR STUDY: DATA, HYPOTHESES, AND ANALYSES

In their monograph Meyers et al. (1964) state that their purpose is "to explore for the presence of a factorial structure in abilities of children of preschool age", and their general hypothesis was that "at all the preschool ages investigated (i.e. 2, 4, and 6 years old), some factor differentiation has occurred" (p.7). Their way to tackle this problem was "to hypothesize four group factors and to build suitable instruments and tests for them". In addition they hypothesized increasing differentiation with increasing age, which should lead to a more detailed factorial structure at later ages, and should allow for decreasing correlations between oblique factors for the older children. Finally they put forward -tentatively- that there should be a greater factor differentiation in normal than in retarded children of the same mental age.

Table 12.1 gives an overview of tests used for the three age groups (2, 4, and 6 year olds), and detailed descriptions can be found in the original publication (p. 9-16). The normal children (85, 89, and 100 for the age groups respectively) were all "Anglo-White" Californians, and the retarded children (56, 40, and 46 for the age groups respectively) were selected from institutions primarily for their testability and for falling within desired mental age brackets (p.19). Note that the designation "two years old" for the retarded children refers only to their *mental* ages (MA), and not their *chronological* ages (CA). The chronological age for the two-year old retarded children ranged from 49 through 175 months, for the four-year olds from 81 through 211 months, and for the six-year olds from 118-214 months.

For each group the Pearson product-moment correlations of the twelve tests were determined from the raw scores (for these correlations see Meyers et al., 1964, p.24,25), the correlation matrices were subjected to a principal component analysis, and subsequently rotated with the biquartimin procedure of Carroll (1957), and with a procrustes rotation to a target loading matrix (Hurley & Cattell, 1962).

We performed a three-mode principal component analysis using the TUCKALS2 program on the six correlation matrices, which will be

Table 12.1 *Four ability-factor study: hypotheses and test names*

Two Years	Four Years	Six Years
Hypothesis A - Hand-Eye Psychomotor		
2-1 Bead stringing (large beads)	4-1 Bead stringing (small beads)	6-1 Bead stringing (same as 4-1)
2-2 Disk stacking (same as 4-2)	4-2 Disk stacking (same as 2-2)	6-2 PMA motor
2-3 Cube stacking (same as 4-3)	4-3 Cube stacking (same as 2-3)	6-3 Circle dotting
Hypothesis B - Perceptual Speed		
2-4 Form-color-size matching	4-4 Pacific color-form matching	6-4 PMA picture matching
2-5 Form-color matching	4-5 Pacific figure matching	6-5 PMA figure matching
2-6 Form matching	4-6 Pacific design discrimination	6-6 Pacific form matching
Hypothesis C - Linguistic Ability		
2-7 Pacific expressive vocabulary and expressive language check list	4-7 Pacific expressive vocabulary (objects and pictures continuous with 2-7 and 6-7)	6-7 Pacific expressive vocabulary (pictures only)
2-8 Pacific receptive vocabulary and receptive language check list	4-8 Pacific receptive vocabulary with Ammons FRPV (continuous with 2-8 6-8)	6-8 Pacific receptive vocabulary with Ammons FRPV (continuous with 4-8)
2-9 Pacific identification by-use	4-9 Response to pictures and Monroe ideational fluency	6-9 Monroe ideational fluency
Hypothesis D - Figural Reasoning		
2-10 Pacific pattern completion	4-10 Pacific object classification	6-10 IPAT classification
2-11 Pacific form and picture comple-	4-11 Pre-Raven pattern completion	6-11 Raven matrices
2-12 Design copying (same as 4-12)	4-12 Design copying and Pacific pattern copying (continuous with 2-12 6-12)	6-12 Pacific pattern copying and design copying (continuous with 4-12)

Source: Meyers et al. (1964). p.14.

labelled N2, N4, N6, R2, R4, R6 (N = normal; R = retarded; i = age) and we also used BMDP4M (Dixon, 1981) to obtain separate principal component analyses (Method = PCA) for each of the correlation matrices. Even though Meyers et al. used a fore-runner of BMDP4M, i.e. BIMED17, we were not able to reproduce their loadings exactly; nor did a principal factor analysis (Method = PFA), be it that the latter results were somewhat closer. In the sequel we will use the PCA results from our analysis with BMDP4M rather than those of Meyers et al.

12.5 FOUR ABILITY-FACTOR STUDY: THREE-MODE ANALYSIS

Test components. In order to follow Meyers et al.'s analysis as closely as possible, four components were determined for the 12 tests, the loadings of the common space for all groups together are given in Table 12.2, and in Fig. 12.1 a visual impression is given of the spatial arrangements of the tests in the subspace spanned by the second, third, and fourth components. It is clear that by a non-singular transformation of the component space a set of oblique axes can be found, each of which represents one of the test groups A, B, C and D (see Table 12.1). It is equally clear that not many new insights will be gained by such a procedure.

Table 12.2 *Four ability-factors study: test loadings*

		components				
		general intelligence	A-C	B	D	
		1	2	3	4	
A	1	27	33	-15	-31	
Hand-Eye	2	27	44	-19	-10	
Psychomotor	3	25	45	-30	-19	
B	4	32	3	48	-10	
Perceptual	5	33	2	49	- 2	
Speed	6	32	2	43	- 4	
C	7	30	-41	-18	-18	
Linguistic	8	26	-44	-14	-29	
Ability	9	26	-34	-27	-24	
D	10	28	- 5	- 6	60	
Figural	11	30	- 9	-22	46	
Reasoning	12	30	9	-14	33	
% explained variation		48	11	8	7	74

Note: decimal points omitted

From Table 12.1 and Fig. 12.3 it may be concluded that for two through six year old normal and retarded children a common structure of the tests is present, and that it conforms to the four (obli-

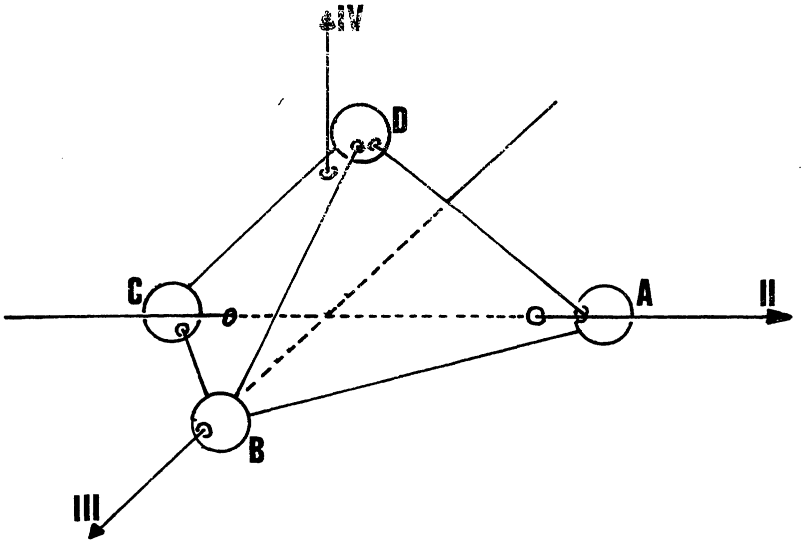


Fig. 12.1 *Four ability-factor: Spatial arrangement of test groups*

(A: Hand-eye Psychomotor; B: Perceptual Speed; C: Linguistic Ability; D: Figural Reasoning)

que) factors hypothesized by Meyers et al. It is equally acceptable to describe the structure as consisting of a general ability (intelligence) component, and a three-dimensional subspace in which the tests form a tetrahedron. The eigenvalues of the second, third, and fourth components are of the same order of magnitude, and the orientation of the axes is thus relatively arbitrary. As we shall see below, the six groups vary with respect to the order of importance of these three components.

Assessing differentiation via the core matrix. From the common component loadings no detailed statements can be made about increasing differentiation with increasing age, or the differences between retarded and normal children. How relevant the components are for each group, can be seen from the extended core matrix (Table 12.3). At this point we only discuss the diagonal elements of the core planes, and we neglect the possible interactions between the com-

Table 12.3 *Four ability-factor study: differences between groups*

A: *T2 scaled diagonal core elements* (approximate percentages explained variation)

mental age		normal			retarded			over-all
		2	4	6	2	4	6	
general intelligence	1	59	37	43	35	59	51	48
A vs C	2	4	11	11	14	9	14	11
B	3	8	7	7	12	6	3	8
D	4	5	6	7	6	4	8	7
sum		76	61	68	67	78	76	

B: *Separate analyses* (percentages explained variation)

mental age		normal			retarded		
		2	4	6	2	4	6
	1	60	38	43	35	59	51
	2	9	12	15	15	10	16
	3	6	9	10	12	7	8
	4	5	7	6	7	6	7
sum		80	66	74	69	82	82

ponents in specific groups. The off-diagonal elements are small, and never larger than the corresponding diagonal elements. The complete core matrix can, by the way, be found in section 5.5, in which these data were used to illustrate procedures for diagonalization of the extended core matrix. For comparison we have included the proportions explained variation of the separate principal component analysis for each group in Table 12.3.

The first thing to notice is that the three-mode analysis provides a fair representation for the structure in each group. The differences in amount of explained variation between the joint and the separate analysis of a group is between four to six percent. In other words, the separate analyses never succeeded in explaning more than six percent over and above the joint analysis. Even the importance attached to the various components tends to be the same in the joint analysis and the separate analyses. Note that the weights or saliences in the three-mode analysis always refer to the same axes (the rows of Table 12.3A), while in the separate analyses the axes may and do have different orientations. To illustrate the latter point the planes spanned by the second and third components for each of the separate analyses are presented in Fig. 12.2.

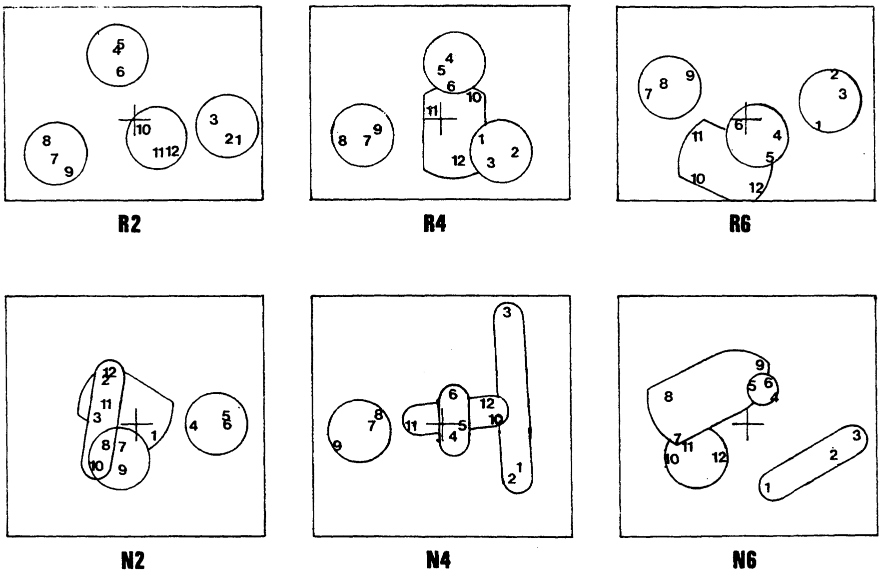


Fig. 12.2 *Four ability-factor study: Subtests spaces from separate analyses.*

(Second versus third components; 1,2,3: Hand-eye Psychomotor; 4,5,6: Perceptual Speed; 7,8,9: Linguistic Ability; 10,11,12: Figural Reasoning; details see Table 12.1)

This figure also illustrates the difficulty of comparing solutions, and of performing rotations to search for similarities between solutions.

The problem with finding suitable rotations is that, except for target rotations, there is no guarantee that even if the same structure is present this structure will actually emerge from the rotated solutions. The results from the biquartimin procedures in Meyers et al. are a case in point. Target rotations are useful in as far as one knows a priori what the structure should be, which was the case for Meyers et al. . Otherwise one could take one of the solutions as target, but then the problem arises which is the best for this purpose.

On the basis of Table 12.3 the question whether there is an increasing component differentiation can be answered. It is useful to discuss the question separately for normal and retarded children. Furthermore, as also remarked by Meyers et al., this question can only be answered within the limitations of this study, one of which is that four groups of abilities were tested, and secondly that the tests for the various ages were not the same but adapted to the specific age level. This introduces some unknowable test-age interactions.

Differentiation for normal children. Keeping this in mind, the impression is that for the *normal* children differentiation of abilities as measured by the tests occurred between ages two and four, and no further differentiation occurred between ages four and six. This conclusion is based on the 59 percent explained variation by the first component for the two-year olds, and the 37 and 46 percent for the older children. Furthermore, the distinction between linguistic abilities (C) and hand-eye psychomotor (A) on component 2 is not present for the two-year olds, but is for the older children. Note that the distinctness or coherence of perceptual speed tests (B) and figural reasoning tests (D) is the same for all age groups.

Meyers et al. did not reach the same conclusion from their analyses (p.46,47). In our opinion this is mainly due to their pre-occupation with normal-retarded comparisons at the same age levels. Furthermore they disregarded the information in the component weights or eigenvalues, and concentrated solely on loadings.

Differentiation for retarded children. The situation for *retarded* children is quite different from that of the normal ones, but one has to keep in mind that the chronological ages of the retarded children are far higher than those of the normal children. Their comparability is, in fact, an assumption by Meyers et al., which need not be true or be the same at each age level. For instance, the differentiation for the retarded two-year olds is quite as large as that for any other group, possibly suggesting that with respect to differentiation the retarded two-year olds are not comparable to normal two-year olds. On the other hand, the situation is reversed for the retarded four and six year old children. They show *less* differentiation than their normal counterparts. A possible explanation might be that differentiation is a different phenomenon for retarded than for normal children. The retarded four and six year old children confirm, by the way, Meyers et al.'s hypothesis that retarded children show less differentiation than normal children.

Again we do not reach exactly the same conclusions as Meyers et al., who state that they cannot find any differences in factor differentiation. As before they only looked at (rotated) loadings, and disregarded amounts of explained variation.

12.6 CONCLUSION

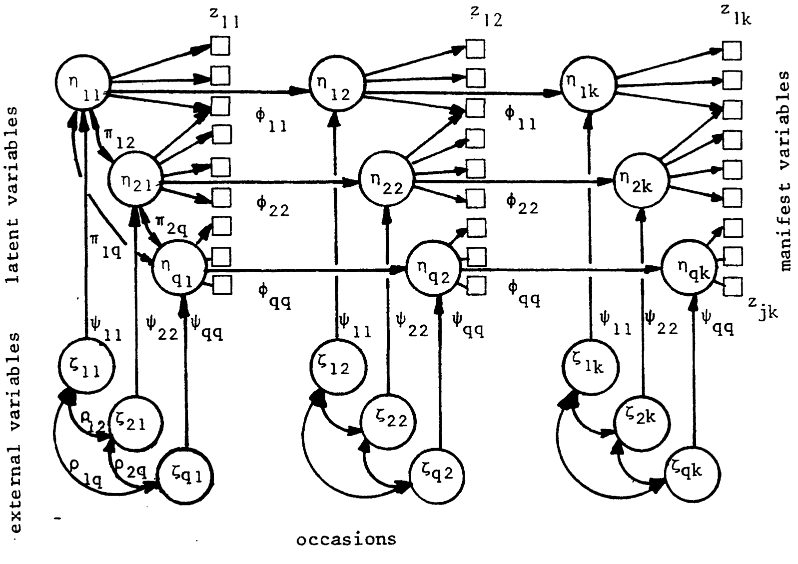
From our analyses it follows that most of the hypotheses of Meyers et al. received some (differentiation) or considerable (four factor structure) support. The difficulty for Meyers et al. was that their level of condensation was simply not high enough. For each correlation matrix with 66 data points they looked at 4 to 5 components, i.e. 48 to 60 loadings, or at 288 to 360 parameters for 396 data points. In our three-mode analysis, the information on which we based our conclusions was contained in 48 loadings and 24 core elements, or 72 parameters in all. This lack of condensation is not really Meyers et al.'s fault. The art was not as developed as it is now, and techniques for dealing with their data in a unified fashion were still being developed.

The gain of using three-mode principal component analysis in comparison with separate analyses and analytic rotation procedures can be considerable. Whether in this particular case even more insight can be obtained via a covariance structure approach is a matter for further investigation.

MULTIVARIATE LONGITUDINAL DATA

13

hospital study



η : latent variables (endogenous variables) - $q=1,\ldots,s$
ζ : external variables (exogenous variables) - $q=1,\ldots,s$
z : manifest variables - $j=1,\ldots,m$
Φ : transition matrix
Ψ : influence of external variables - independent of k
P : correlations of external variables - independent of k ; $\{\rho_{qq'}\}$
Π : correlations of latent variables
occasions - $k=1,\ldots,n$

13.1 INTRODUCTION

In this and the following chapter we turn to the analysis of multivariate longitudinal data. Such data introduce a number of complications not present in 'standard' three-mode data. Whereas in the standard case the main dependence is between variables, in longitudinal data the serial dependence or autocorrelation between observations on different occasions is important as well. The interactions between the two kinds of dependence introduce further complications.

Here we do not attempt to give an in-depth treatment of the subject, but we will try to show that three-mode principal component analysis can assist in making meaningful statements about multivariate longitudinal data.

For the analysis of such data two major approaches may be distinguished. The first approach deals with data from a design in which there are few or no replications, and in which many observations in time are needed to make meaningful statements about the observational units. The time series produced are, generally, analysed by fitting specific stochastic models, such as autoregressive and/or moving average models (ARIMA-models). Especially in the field of econometrics this approach is taken to describe and predict the behaviour of national economies or individual firms. An introduction to this approach can be found in Chatfield (1975), or with a view to applications in the social sciences in Glass, Wilson, & Gottman (1975).

In the second approach one usually has a design with many variables, many observational units, and rather fewer points in time. The main interest with such designs focuses on analysing correlational or covariance structures of the variables at each occasion, between occasions, or for all occasions simultaneously. Wohlwill (1973, p.240) goes as far as stating that "a good case could be made for the proposition that correlational analysis, however denigrated in certain quarters, is the method par excellence for developmental study." On the other hand, Anderson (1963) considers factor analysis (i.e. the correlational approach) only really useful in reducing large numbers of variables to a small number of factors (similar to forming indicators in econometrics), the scores on which can then be subjected to a time series analysis. However valid Anderson's point of view is in the presence of large numbers of repeated measurements on the same observational units, for many social science data the number of occasions is too small to consider fitting the kind of functions Anderson has in mind. In addition, the amount of variation accounted for by the factors will in many cases be too small for reliable estimation of the factor scores anyway. Finally the difference in emphasis remains: in the first approach the stress is on modelling serial dependence for each of the variables, in the second approach on modelling the interrelationships of the variables over time.

Within the correlational approach the distinction between the 'classical' approach to the analysis of cross-product (or correlation) matrices treating all modes fixed, and the 'statistical' approach using the theory of covariance structures treating one mode stochastically, is particularly relevant (see section 3.6 for a discussion of this distinction). In fact, in situations with sufficient knowledge of substantive theory, given enough observational units and no interest in individual differences, the covariance structure approach as advocated and developed by Jöreskog (1978, 1979), Jöreskog & Sörbom (1977), and Bentler (1978,1980), or three-mode path analysis (Lohmöller & Wold, 1980,1982), seem ideal ways to proceed. Lacking the prerequisites, i.e. sufficient substantive theory and observations, one has to make do with less powerful methods which approach longitudinal data in a more exploratory fashion, such as common factor analysis and principal component analysis. Visser (1982), and especially Bentler (1973), discuss the many proposals in this field, and Hakstian (1973) and Corballis (1973) also provide valuable contributions.

13.2 SCOPE OF THREE-MODE ANALYSIS FOR LONGITUDINAL DATA

Wohlwill (1973, p.275-283) and Bentler (1973, p.161-162) mention briefly that three-mode factor analysis and principal component analysis have some potential for treating multivariate longitudinal data, but both authors indicate that very little experience with these techniques is available, and find the real potentialities therefore difficult to assess.

The promise of three-mode principal component analysis and its analogues in the covariance structure approach lies in the simultaneous treatment of serial and variable dependence. The serial dependence can be assessed from the component analysis of the time mode, the variable dependence from the variable mode, and their interaction from the core matrix or the latent covariation matrix (see section 13.3). By analysing variables over subjects ànd occasions with standard principal component analysis (e.g. Vavra, 1972; Visser, 1982, p.63,172), or by treating variables at each occasion as separate variables, and analysing these with standard principal component analysis (e.g. Visser, 1982, p.64, 151ff,172), the variable and serial dependence, and their interactions may become confounded.

Multivariate longitudinal data have been analysed with three-mode principal component analysis or its covariance structure analogues by Tucker (1967), Love & Tucker (1970), Inn, Hulin, & Tucker (1972), Lammers (1974), Van de Geer (1974), Lohmöller (1978, 1981a), Hanke, Lohmöller & Mandl (1980), Snyder, Bridgman, & Law (1981), Haan (1981), Skolnick (1981), and Harshman & Berenbaum (1981).

Lohmöller (1978, 1981a) has made some progress towards the interpretation of serial and variable dependence in three-mode analysis; his contribution will be discussed in the next section.

13.3 ANALYSIS OF DATA FROM MULTIVARIATE AUTOREGRESSIVE PROCESSES

Introduction. In this section we will review, discuss and amend proposals by Lohmöller (1978, 1981a) to interpret the results from a three-mode principal component analysis of longitudinal data. Basic to his approach is the assumption that the changes in the variables and scores of the subjects on these variables can be modelled by multivariate autoregressive processes. Given that the assumption is tenable, Lohmöller shows how one can get an indication of the size of the parameters of the assumed autoregressive process from the results of the three-mode analysis. In other words, once a three-mode analysis has been performed, an interpretation of its results can be given in terms of the parameters of the autoregressive process.

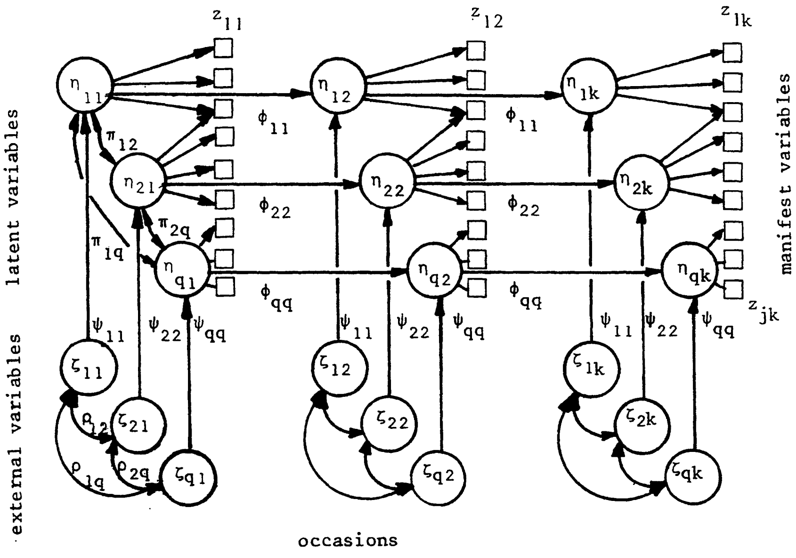
Lohmöller's procedure to arrive at these indicators or 'estimators' is rather indirect. Some 625 data sets were first generated according to specific autoregressive processes, and then analysed with three-mode principal component analysis. Empirical 'estimating' equations were derived for the parameters of the autoregression parameters by regressing these parameters on parameters in the three-mode analyses. Lohmöller himself recognized that the procedure has serious drawbacks, as for each new kind or size of data set new simulation studies have to be performed. On the other hand, via his simulation studies Lohmöller was able to investigate which results in a three-mode analysis are particularly sensitive to changes in specific parameters in the autoregressive models, and which results reflect the general characteristics of the autoregressive models. Lohmöller (1981a, p.70) pointed out that, in fact, three-mode path models (see section 3.8) are to be preferred, because in those models the autoregressive processes can be modelled directly. The example in section 13.4 is, however, too large to be handled by such an approach, as it would require the analysis of a 242×242 covariance matrix.

Component analysis of time modes. One of the problems in three-mode analysis of longitudinal data is the interpretation of the decomposition of the time mode, as its correlation matrix is more often than not a (quasi-) simplex. Entries of a simplex are such that the correlations between two time points are a decreasing function of their difference in rank order. Typically the bottom-left and top-right corner of the correlation matrix have the lowest entries (for examples see Table 13.2). As Guttman (1954) has shown, simplexes have standard principal components. In particular, the components of equidistant simplexes can be rotated in such a way that the loadings on the first component are equal, those on the second component are a linear function of the ranks of the time points, those on the third component are a quadratic function of the ranks, etc. After extraction of the first two principal components all variables have roughly the same communalities, and the configuration of time points resembles a horseshoe or "C", the opening of which is increasing with the relative size of the first eigenvalue (see Borg, 1976; see section 2.7 for an example of a horseshoe).

The problem is not so much the standard solution as the fact that there are many different processes which could have generated the simplex (for an amusing example see Fischer, 1967). It is, therefore, next to impossible to interpret the results of a component analysis of a time mode without a substantive theory about the underlying processes which generated the data.

One kind of change which produces correlation matrices with a (Markov) simplex structure (see e.g. Jöreskog, 1970,1974; and Morrison, 1976, section 9.11) are first-order autoregressive processes, to be discussed below. Thus a simplex-like correlation matrix *may* be explained by such an autoregressive process. Furthermore, a process of steady growth with a level and a gain component produces a correlation matrix with a (Wiener) simplex form (see Jöreskog, 1970,1974, and Morrison, 1976, section 9.11). Thus one *may* describe the underlying change process, at least approximately, by a level and a gain component. In the case that one has a simplex-like correlation matrix, the results of a component analysis on this correlation matrix *may*, therefore, be interpreted both in terms of parameters of the autoregressive process, and in terms of a level and a gain component, provided the subject matter warrants the use of such models.

Autoregressive processes. In Fig. 13.1 an example is given of the kind of autoregressive models one might consider. The model was suggested by Lohmöller (1978) in the context (and we will follow his description) and presented in some detail by Roskam (1976, p. 126-130), who also used it to model change processes, and discussed the merits of this and similar models.



η : latent variables (endogenous variables) - q=1,...,s
ζ : external variables (exogenous variables) - q=1,...,s
z : manifest variables - j=1,...,m
Φ : transition matrix
Ψ : influence of external variables - independent of k
P : correlations of external variables - independent of k ; $\{\rho_{qq'}\}$
Π : correlations of latent variables
occasions - k=1,...,n

Fig. 13.1 *A multivariate autoregressive model*

The structural part of the model (see also section 3.8) has the form of a multivariate regression equation; it is assumed that the state of the latent variables $\underline{\eta}_k$ at occation k depends on only two influences, viz. the state of the variables at occasion k (i.e. it is a first-order process), and the state of the external variables $\underline{\zeta}_k$ at occasion k:

$$\underline{\eta}_k = \Phi\underline{\eta}_{k-1} + \underline{\Psi}_k \qquad k=1,\ldots,n$$

where $\underset{\sim}{\eta}_k$ and $\underset{\sim}{\eta}_{k-1}$ are the vectors of all t latent variables. In the model it is assumed that the external variables $\underset{\sim}{\zeta}_k$ and $\underset{\sim}{\zeta}_{k'}$ ($k \neq k'$) are uncorrelated, which is almost always an oversimplification as it implies that all time dependent influences are included in the model. Maybe somewhat incorrectly, we have written the external variables as latent rather than manifest ones. Finally, it is assumed in this model that the latent and external variables are normalized.

The matrix Φ, called the *transition matrix*, describes how strongly the latent variables at occasion k-1 influence those at occasion k. When Φ is diagonal, as is the case for the model in Fig. 13.1, the latent variables only influence themselves, and no other latent variables (i.e. $\phi_{qq'} = 0$, $q \neq q'$). When Φ is diagonal, the changes in the component structure of the variables are entirely due to external influences.

The matrix Ψ describes the influences of external variables on occasion k on the latent variables on occasion k. The external variables represent the entire influence of the environment on the latent variables. When Ψ is diagonal, as in the model on Fig. 13.1, then each external variable only influences one latent variable (i.e. $\psi_{qq'} = 0$, $q \neq q'$). Note that here the matrixes Φ and Ψ are assumed to be independent of k, so the model assumes that the first-order influences remain identical over time. Differences in influence over time of both latent and external variables cannot be accounted for by this particular autoregressive process, and as such it is almost always an oversimplification of reality.

The structural part discussed until now has been entirely in terms of latent variables, and therefore we also need a measurement model to link the data with the structural model. In the present case, the measurement model is simply the three-mode principal component model itself, in which the components are the latent variables of the autoregressive model.

Latent covariation matrix. Before entering into a discussion of the role autoregressive models can play in three-mode analysis, it is necessary to look at what we will call the *latent covariation matrix S*, called 'core covariance matrix' by Lohmöller (1978). The covariation $s_{pq,p'q'}$ are the inner-products of the elements of the

Tucker2 core matrix:

$$s_{pq,p'q'} = \sum_{k=1}^{n} \tilde{c}_{pq,k}\tilde{c}_{p'q';k} \ ,$$

where we assume that the n observational units or subjects constitute the third unreduced mode (k-mode) in a Tucker2 model with a first i-mode of variables, and a second j-mode of occasions of conditions.

The core elements $\tilde{c}_{pqk}$ may be interpreted as scores of the observational units on the s×t *combination components* of the first and second mode. In this context the a-th combination component is an $\ell \times m$ vector f_a with elements f^a_{ij}, and $f^a_{ij} = g_{ip}h_{jq}$, $ij=1,\ldots,\ell m$, and $a=1,\ldots,st$. In the example of section 13.4 one of the combination components is, for instance, labelled as 'gain (=trend or time component) in degree of specialization (= latent variable or variable component)'. In that case a $\tilde{c}_{pqk}$ represents the score of the k-th hospital on the combination component 'gain in degree of specialization'.

The value of s_{aa} (= $s_{pq,p'q'}$) thus indicates the covariation of the a-th (=pq-th) and a'-th (=p'q'-th) combination components. Within the analysis of covariance structures (see section 3.5), in which the mode of observational units is stochastic, the latent covariation matrix arises in a natural way. If we define $\underline{y}_a$ ($a=1,\ldots,st$) to be the random vector of scores on the combination components, then the three-mode model can be written as

$$\underline{z}_{ij} = \sum_{a=1}^{st} f_{ij,a}\underline{y}_a = \sum_{p=1}^{s}\sum_{q=1}^{t} g_{ip}h_{jq}\underline{y}_{pq} \qquad \text{with } ij=1,\ldots,\ell m$$

and the covariance matrix $\underline{\Sigma} = E(\underline{z}\underline{z}')$ is

$$\underline{\sigma}_{ij,ij} = \sum_{p=1}^{s}\sum_{p'=1}^{s}\sum_{q=1}^{t}\sum_{q'=1}^{t} g_{ip}g_{i'p}\, h_{jq}h_{j'q'}\, \underline{s}_{pq,p'q'} =$$

$$= \sum_{a=1}^{st}\sum_{a'=1}^{st} f_{ij,a}f_{i'j',a'}\underline{s}_{aa'}$$

with $\underline{S} = E(\underline{y}\underline{y}') = \{\underline{s}_{pq,p'q'}\} = \{\underline{s}_{aa'}\}$ the latent covariation matrix. Loosely speaking one may say that the latent covariation matrix underlies the observed covariation matrix, and embodies the basic covariations present in the data. Lohmöller scaled the latent covariation matrix, analoguous to Bartussek's (1973) method of scaling the core matrix (see section 6.9); we have also done so in the example in section 13.4 for comparability with his results (see Table 13.5).

It is, by the way, also possible to develop the latent covariation matrix from a Tucker3 model, but scores on combination components must then be interpreted as produced by 'idealized subjects' rather than real ones. This, at times, might be less convenient than referring to the subjects themselves.

The major purpose in discussing the latent covariation matrix from a Tucker2 model is that it provides a means of investigating the relationships between autoregressive processes and three-mode analysis. It is especially the structure of the latent covariation matrix which can be used to investigate the parameters of the postulated autoregressive process underlying the observations. Lohmöller (1978, 1981a) uses the latent covariations to derive empirical estimation equations for the parameters of the autoregressive parameters. We will not follow him in directly applying these equations to our example in section 13.4, as it does not conform to the restriction built into Lohmöller's simulation studies. On the other hand, from these studies Lohmöller derived more general relationships between the latent covariations and general characteristics of the autoregressive processes, which seem to be applicable outside the specific design of his simulation studies.

It is equally possible to interpret the variations and covariations directly. The latent variations, s_{aa}, of the combination components may be divided by the total variation present in the data - SS(Tot)-, and can then be interpreted as the porportion variation explained by the combination component in question. After all

$$s_{aa} = s_{pq,pq} = \sum_{k=1}^{n} \tilde{c}^2_{pqk} ,$$

and in section 6.9 we showed that the squared elements of the core matrix can be interpreted as explained variation. The covariations $s_{aa'}$ can be transformed into direction cosines between the combination components a and a':

$$s^{*}_{aa'} = s_{aa'}/s^{\frac{1}{2}}_{aa}s^{\frac{1}{2}}_{a'a'} \quad (a=1,\ldots,st).$$

In section 6.9 we discussed a similar interpretation for the components themselves when the first and second mode elements refer to the same quantities. Interpreting latent covariations in this way is more direct, and has a wider applicability than the interpretation via parameters of autoregressive processes. On the other hand, the latter interpretation gives more specific and more substantive information because of the postulated model.

Linking autoregressive parameters to three-mode results. In the introduction we referred to two major sources of dependence in multivariate longitudinal data: variable and serial dependence. It is of interest to know if, and if so, in which way these kinds of dependence influence each other. One may have a structure between variables (variable dependence), which is *stationary*, i.e. not changing in time. It is also possible that the subjects maintain their relative positions on the variables, irrespective of the structure of the variables, i.e. there is *stability* of the variables or high autocorrelations (serial dependence). Finally, variables and subjects may change simultaneously in which case there is stability nor stationarity.

In the above paragraph one may also read 'latent variables' instead of 'variables'. In general we will discuss latent variables as the autoregressive processes were formulated in those terms.

So, we will refer to a set of latent variables as being *stationary* when the same component structure is present on all occasions. In particular we will call the set *homogeneous* when the variables are highly correlated, and can be represented in a low-dimensional space, i.e. by a few components. In autoregressive processes the homogeneity is indicated by the covariations of the latent variables at each time point (the π_{qqk} in Fig. 13.1).

In three-mode analysis we derive one set of orthonormal variable components over all occasions simultaneously, and the dimensionality of the component space is thus an indication of the overall homogeneity. As there is only one component matrix for the variables, one could get the impression that the model does not allow for non-stationarity. This is, for instance, indeed the case in a model without a core matrix like PARAFAC1 (see Harshman & Berenbaum, 1981), but not in the Tucker2 model, in which the deviations from stationarity show up in the core matrix and the latent covariation matrix. Lohmöller's contribution is that he attempted to investigate how and what kind of stationarity could be inferred from the latent covariation matrix. He claims that for an autoregressive process as shown in Fig. 13.1, increasing and decreasing homogeneity, ceteris paribus, can be gleaned from the size and the signs of the covariations between the latent variables in the latent covariation matrix. We will return to this point in some detail when discussing the example in section 13.5.

A (latent) variable will be called *stable* when the relative positions of the observational units on that (latent) variable stay the same in time, and the stability of a (latent) variable may be judged from the covariations of the latent variables on different occasions. A set of variables will be called stable when all variables are stable. In autoregressive processes the stability of a latent variable, $\underline{\eta}_q$, is given by ϕ_{qq}; ϕ^2_{qq} indicates to what extent the latent variable is determined by its predecessor. Stable variables are sometimes called trait-like, i.e. mainly determined by the defining construct or trait, and unstable variables are sometimes called state-like, i.e. mainly determined by the time they are measured (see Cattell, 1966b, p.357). The size of covariations of a latent variable between time points is thus an indication for the stability of a variable with high values indicating a trait-like and low values a state-like variable. The overall stability of a set of variables, $\bar{\phi}$, may be determined from the first eigenvalues of the time mode, given that first-order autoregressive processes underlie the data (Lohmöller, 1981a, p.29). When a second-order model holds, the stability will most likely be overestimated using the first eigenvalue. A global check on the appropriateness of the 'estimate' may be made by comparing its value to the lag-one correlations.

Lohmöller also investigated the structure of the latent covariation matrix for autoregressive models like those in Fig.13.1 in case of different stabilities of the latent variables, and in case of equal stabilities on changing dimensions. We will, however, not go into that part of his study.

Finally, it is interesting to consider the situation in which all latent variables have equal stability (i.e. uniform autocorrelations) the set of latent variables is stationary, and no partial cross-lag correlations exist. For this situation Lohmöller (1978, p.4) showed that the latent covariation matrix is an (st×st) identity or diagonal matrix, depending on the particular scaling of the components. This means that the three-mode model for the observed covariation matrix reduces to

$$\sigma_{ij,i'j'} = \sum_{p=1}^{s} \sum_{q=1}^{t} g_{ip} g_{i'p} h_{jq} h_{j'q} v_{pq,pq} = \sum_{a=1}^{st} f_{ij,a} f_{i'j',a} v_{aa} \quad ,$$

with $v_{pq,pq} = v_{aa}$ equal to one or non-zero depending on the scaling of the components. This model may be used as a kind of 'null-hypothesis' to evaluate latent covariation matrices (see also section 3.5).

Discussion. The above approach to evaluating change phenomena in multivariate longitudinal data depends very much on the appropriateness of the multivariate autoregressive models. It is also still very sketchy from a mathematical point of view, and therefore requires further investigation. Further practical experience is also necessary to assess its potential. It seems that in some cases the assumption of an underlying autoregressive process is not unreasonable, as in the example in section 13.4. In connection with this example we will discuss rough and ready ways to assess whether the assumption of the autoreggressive model is tenable.

Lohmöller's major contribution is that he provides a framework for the interpretation of multivariate longitudinal data, which cannot easily be handled directly by causal modelling or time series analysis. In addition, he gives yet another possibility of interpreting the core matrix (and the latent covariation matrix) (see also section 6.9).

13.4 GROWTH AND DEVELOPMENT OF DUTCH HOSPITAL ORGANIZATIONS

Research questions. In order to gain some insight into the growth and development of large organizations, Lammers (1974) collected data on 22 organizational characteristics (variable or j-mode) of 188 hospitals in the Netherlands ('subject' or i-mode) from the annual reports of 1956-1966 (time or k-mode). His main questions with respect to these data were:

1. whether the organizational structure as defined by the 22 variables was changing in time;
2. whether there were different kinds of hospitals with different organizational structures and/or different trends in their structures.

These two questions will be taken up in this section.

In the next section we will also look at such questions as:

3. do the latent variables have different stabilities;
4. is the latent variable domain stationary?
5. is an interaction present between serial and variable dependence, i.e. do the latent variables and the trends interact;

In other words, in section 13.5 we will try to assess the parameters of a possibly underlying autoregressive process.

Data. Prior to the three-mode analysis, the majority of the variables were categorized into roughly ten intervals of increasing length for practical reasons (removing skewness of the counted variables, easing visual inspection, preparing the data for other analyses, etc.). The variables were normalized over all year-hospital combinations (i.e. j-normalization, see section 6.2), thus removing incomparable means and standard deviations, while maintaining the trends over the years. In Table 13.1 the variables are given with their categorizations and mnemonics.

Results of three-mode analysis. To answer the first question with respect to the changes in organizational structure it is necessary to examine first of all the structure itself. We will do this by inspecting the joint plot (see section 6.10) of the hospitals and variables using the first trend core plane (Fig. 13.2).

Table 13.1 *Hospital study: variables, their mnenomics, and categorizations*

nr	mnemonic	variable	categories
1	TRAI	*training capacity*	number of training facilities
2	RESC	*research capacity*	1: no research or experiments 2: radio-active isotope research or animal experiments 3: radio-active isotope research *and* animal research
3	FIND	*financial director*	present or absent
4	FACI	*facility index*	number of facilities such as laboratories and libraries
5	WARD	*ratio qualified nurses in outside wards*	1:0.00-0.99 5:4.00-4.99 8:7.00-7.99 2:1.00-1.99 6:5.00-5.99 9:8.00-8.99 3:2.00-2.99 7:6.00-6.99 10:none outside wards 4:3.00-3.99
6	QUAN	*ratio qualified nurses/ total number of nurses*	1:0.01-0.30 3:0.41-0.50 5:0.61-0.70 2:0.31-0.40 4:0.51-0.60 6: > 0.70
7	FUNC	*number of functions*	1: 1-10 3:16-20 5:26-30 7: > 35 2:11-20 4:21-25 6:31-35
8	STAFF	*total staff*	1: 1- 50 5:201-250 9:401-450 12:551-650 2: 51-100 6:251-300 10:451-500 13:651-750 3:101-150 7:301-350 11:501-550 14: 750 4:151-200 8:351-400
9	RUSH	*Rushing index*	spread of work: $RUSH = 1 - \frac{\Sigma x^2 \ / \ (\Sigma x)^2}{1 - (1/N)}$ (x = number of people having a function, N = number of functions) 1:.00<R<.80 4:.84≦R<.86 6:.88≦R<.90 2:.80≦R<.82 5:.86≦R<.88 7:≧ .90
10	EXEC	*executive (managerial and supervising) staff*	1: 1- 5 3:11-15 5:21-25 7:31-35 9:>40 2: 6-10 4:16-20 6:26-30 8:36-40
11	NMPR	*non-medical professionals*	number of pharmacists, psychologists, etc.
12	ADMI	*administrative (i.e. clerical) staff*	1: 0 3: 6- 10 5: 16- 20 7:> 30 2: 1- 5 4: 11- 15 6: 21- 30
13	PARA	*paramedical staff*	1: 0 4: 11- 15 7: 26- 30 10: 51- 60 2: 1- 5 5: 16- 20 8: 31- 40 11:>60 3: 6-10 6: 21- 25 9: 41- 50
14	NMED	*other non-medical staff*	1: 1-10 3: 31- 50 5: 71- 90 7:111-150 2: 11-30 4: 51- 70 6: 91-110 8:>150
15	NURS	*total number of nurses*	1: 1-25 4: 76-100 7:151-175 10:>300 2: 26-50 5:101-125 8:176-200 3: 51-75 6:126-150 9:201-300
16	BEDS	*total number of beds*	1: 1-50 4:151-200 7:301-400 2: 51-100 5:201-250 8:401-600 3:101-150 6:251-300 9:>600
17	PATI	*total number of patients*	1: 1-1000 5:4001-5000 8:7001-8000 2:1001-2000 6:5001-6000 9:8001-9000 3:2001-3000 7:6001-7000 10:> 9000 4:3001-4000
18	OPEN	*openness*	closed/partly closed/open to consulting physians from 'National Health' and private patients
19	CMSP	*main clinical specializations*	number of specializations
20	PMSP	*main polyclin. specializations*	number of specializations
21	CSUB	*clinical subspecializations*	number of specializations
22	PSUB	*polyclin. subspecializations*	number of specializations

This trend or time component contains most of the variation in the time domain, and reflects the overall characteristics of the components of the variable and hospital domains (see also below).

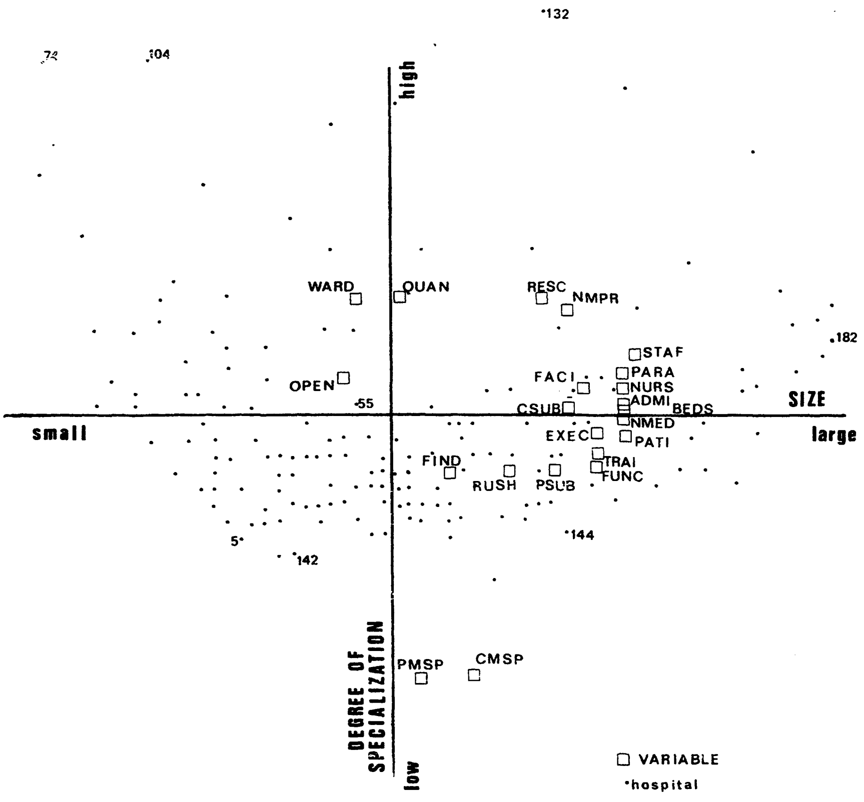


Fig. 13.2 *Hospital study: Joint plot of hospitals and variables*

The joint plot shows that the latent variables may be interpreted as *size* and *degree of specialization*, where the former is not only indicated by the variables intended to measure size (see Table 13.1), but also by most of the other variables. The degree of specialization is primarily indicated by a deficit of main specializations (PMAIN,CMAIN), a larger research capacity (RESC), and greater proportions of qualified nurses (QUAN) and qualified nurses on the wards (WARD). The hospital components or prototype hospitals will be designated as *general hospitals* and *specialized hospitals* respectively. From the relative sizes of the standardized component weights ($\lambda_1 \cong \mu_1 \cong .50$; $\lambda_2 \cong \mu_2 \cong .06$) we may conclude that the first components are by far the most important ones. The second components essentially arise from the fact that some 15-20 hospitals have sizeable deficits in main specializations compared to the other hospitals. Incidentally, the sharp boundary of the hospitals on the negative Y-axis in Fig. 13.2 is caused by ceiling effects due to the fact that a large number of hospitals have all the main specializations a hospital can have.

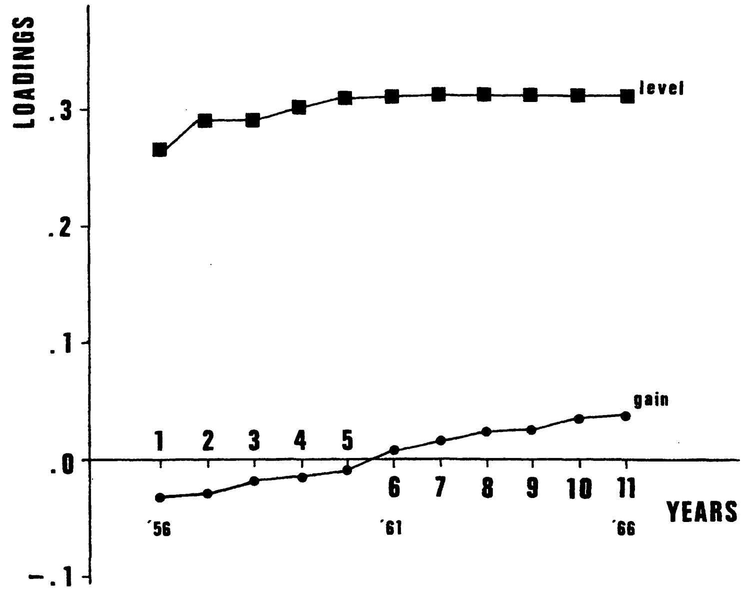


Fig. 13.3 *Hospital study: trends* (scaled)

For the inspection of the time mode components (Fig. 13.3.) it is advantageous to scale the components in accordance with their eigenvalues ($v_1 = .55$; $v_2 = .006$), as an assessment of their relative importance is crucial (see also section 6.8). The figure shows that the overall structural organization remains the same, except

Table 13.2 *Hospital study: correlations between the various years*

time mode
(based on 188 x 22 observations)

	1	2	3	4	5	6	7	8	9	10	11
1	100										
2	96	100									
3	94	97	100								
4	93	95	98	100							
5	89	92	94	95	100						
6	87	90	92	93	97	100					
7	86	88	90	91	94	95	100				
8	85	88	90	91	94	95	97	100			
9	83	85	87	89	91	93	94	96	100		
10	81	84	86	87	90	91	93	95	97	100	
11	80	82	85	86	89	90	92	94	95	97	100

number of beds

	1	2	3	4	5	6	7	8	9	10	11
1	100										
2	97	100									
3	97	98	100								
4	96	98	99	100							
5	95	96	98	98	100						
6	94	96	97	98	99	100					
7	94	95	97	97	98	99	100				
8	93	95	96	97	98	98	99	100			
9	92	94	95	95	97	97	98	99	100		
10	92	93	94	94	96	96	96	98	99	100	
11	91	92	93	93	95	95	95	97	97	99	100

main polyclinical specializations

	1	2	3	4	5	6	7	8	9	10	11
1	100										
2	94	100									
3	90	92	100								
4	87	90	98	100							
5	79	82	90	90	100						
6*	70	72	79	80	89	100					
7	72	74	82	83	86	79	100				
8	74	77	84	86	89	83	93	100			
9	70	72	79	80	82	76	88	93	100		
10	66	68	77	78	80	75	86	90	96	100	
11	65	66	75	77	79	72	84	89	95	98	100

* Note the curious break in the simplex by years 6 and 7

for a slight increase in the first years (say, '56-'61), as the first trend shows a very strong stable *level*. The second trend, *gain*, shows a very steady increase, but is relatively unimportant.

From section 13.3 we know that we may expect such components from longitudinal data showing a simplex structure in the time mode. Table 13.2 shows the correlation matrix of the time mode, and of two of the variables. Thus, the answer to the first research question is that the overall organizational structure is stable, i.e. the relative position of the hospitals remained unaltered, but there is a steady but small increase or decrease in overall level or size, depending on the signs of the loadings on other components (see e.g. Fig. 13.4).

Some authors (Van de Geer, 1974; Lohmöller, 1981a) suggest that it is advantageous to rotate the components from a simplex to orthogonal polynomials, leading to a first component which has equal entries, a second component with entries increasing linearly over time when the time points are equidistant, and a third component which shows a quadratic function of time, i.e. first an acceleration and then a deceleration, or vice versa. For the present data, it was attempted to rotate the time mode to such a matrix of orthogonal polynomials, but the rotation matrix was practically an identity matrix (r_{11} = .9996; r_{12} = r_{21} = .0294; r_{22} = .9996). Not surprisingly, it only transferred a very small amount of the growth in overall level from the first component to the second component. We will, therefore, continue to use the unrotated time components.

To answer the second research question a decomposition in terms of components alone does not suffice, and the core matrix must be inspected as well. First of all, Table 13.3 confirms the answer to the first research question. The combination of the first components of all three modes (general hospitals, size, and level), c_{111}, explains most of the fitted variation [SS(Fit) due to c_{111})/ SS(Fit overall) = .49/.56 = .88; see section 6.9]. The gain in size of the general hospitals, c_{112}, is negligible over and above the increase already contained in the level component.

The second important combination (c_{221}=53) indicates that the specialized hospital also maintain their overall level of specialization. There is a slight tendency (c_{222} = -5) to become less specialized, and to grow in overall size (c_{212} = 10). Similarly the large general hospitals tend to become somewhat less specialized (c_{221}= -7). The standardized contributions to the SS(Fit) show that these effects are very small, leading to the conclusion that the specialized hospitals do not have a very different growth pattern from that of the other hospitals.

Table 13.3 *Hospital study: core matrix*

raw core matrix

	general hospitals			specialized hospitals	
	size	degree of specialization		size	degree of specialization
level	150	1	level	-1	53
gain	2	-7	gain	10	-5

standardized contribution to the fitted sum of squares*

level	.49	.000	level	.000	.06
gain	.000	.001	gain	.002	.000

* The standardized contribution of a core element c_{pqr} is c^2_{pqr}/SS(Total). SS(Total) = 45496

A more detailed inspection of the time mode is given in Fig. 13.4, where the elements of the extended core matrix with *years* as unreduced third mode of a TUCKALS2 analysis on the same data (hospital×variables×years as first, second, and third mode respectively) have been plotted against time. The patterns are, of course, in accordance with the TUCKALS3 analysis, but the development of the relations between the hospital and variable components over time is shown more explicitly.

13.5 INTERPRETATION IN TERMS OF AUTOREGRESSIVE PROCESSES

'Estimation' of change phenomena and methods of analysis. When 'estimating' the values of the parameters in autoregressive processes a difference must be made between the 'estimators' derived from

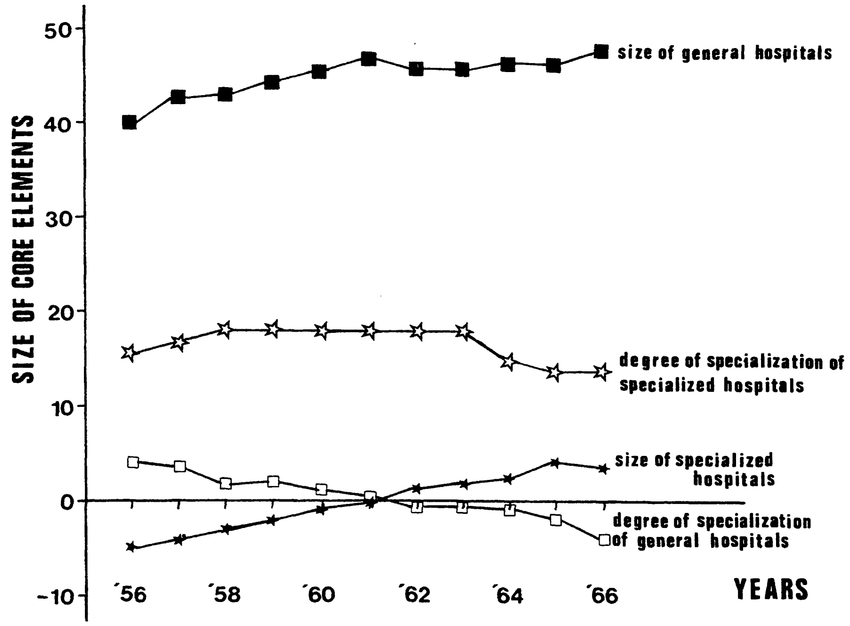


Fig. 13.4 *Hospital study: Trends for prototype-latent variable combinations* (based on extended core matrix)

the classical Tucker approach, and those from the alternating least squares (ALS) approach to solving the three-mode model (see Chapter 4 for a discussion of the two approaches). In the former approach the eigenvalues reported for each of the modes are based on the cross-products of the raw data, while in the latter approach the eigenvalues belong to the data, which have been simultaneously reduced over the two other modes. This implies that in the Tucker approach the amount of variation accounted for by each mode can be different, and larger than the amount of variation accounted for by the simultaneous ALS approach. In the latter case the amount of variation accounted for is more or less the intersection of those of the three modes of the Tucker approach (see also the discussion in section 4.5). As argued in section 3.6 the ALS approach seems more appropriate, exactly because a precise amount of variation can be ascribed to the three-mode solution.

As Löhmoller (1978, 1981a) uses the Tucker method, his detailed estimation equations and his tables should be used with some caution when employing the outcomes from an ALS analysis. Of course, when the differences are small no real problems arise. At first glance it seems best to inspect, for instance, the correlation matrices based on the reconstructed data estimated from the model rather than the raw score correlation matrices themselves, because it is on the model part of the data that the substantive conclusions are based, but this approach has not yet been investigated. Here we will only derive rough and ready indications of some parameters of the multivariate autoregressive models, and not rely on the details of Lohmöller's studies.

Checking the order of autoregressive process. Before attempting to estimate the parameters of an autoregressive process it should be established whether it is reasonable to postulate such a process for the data, and if so, whether it is of the right order. There seem to be a number of ways to do this.

First, check whether the time mode is a simplex, as we know that autoregressive processes generate simplexes. Formal methods are referenced in Morrison (1976; section 9.11). Inspection of the hospital data shows the correlation matrix to be a simplex (see Fig. 13.3), moreover the points in time (years) are equidistant.

Secondly, perform a multiple regression of z_k for the data at time k on z_{k-1}, z_{k-2}, ... (z_k is the ($\ell m \times 1$) vector of for the k-th occasion). When the autoregressive process is a first-order one, only z_{k-1} should have a sizeable regression weight. Although ordinary least squares estimation will lead to incorrect standard errors for the estimators of the regression weights, they are in general unbiased (see e.g. Visser, 1982, p.71). Table 13.4 shows the standardized regression weights, and the hospital data seem to follow at least a second-order autoregressive process with a dominant first order. This implies, that, for instance, the lag-one correlations will underestimate the overall stability of the process.

Combining the above information, the assumption of a second-order autoregressive process with a rather strong first-order seems plausible. However, as pointed out in section 13.3, Lohmöller's procedures were only developed for first-order autoregressive processes, because higher-order autoregressive processes turned out to be unmanagable. For illustration's sake, we will follow him in his proposal and continue as if the process is first-order. In other words, we pretend that the autoregressive process is a first-order one, keeping in mind this is only an approximation.

Table 13.4 *Hospital study: Standardized regression weights for predicting an occassion from earlier occasions*

t	t-1	t-2	t-3	t-4	t-5	t-6	t-7	t-8	t-9	t-10
2	96	-								
3	79	19	-							
4	85	14		-						
5	75	14	7		-					
6	87	10				-				
7	63	19	10				-			
8	70	16	8	5				-		
9	75	13	12				2		-	
10	76	15	4	4						-
11	75	11	8							

-all values have been multiplied by 100.
-analyses were performed with BMDP1R; Dixon, 1981.
-standard F-to-enter of 4.00 was used, and a tolerance of .01.

Assessment of change phenomena in the Hospital study. In this subsection we will apply Lohmöller's proposals for assessing change phenomena to the hospital study. In order to remain compatible with his discussion, we will use the results from the jk-normalized data (i.e. normalized per variable on each occasion), instead of those from the j-normalized data (i.e. normalized per variable over all occasions together). None of the substantive results reached so far are seriously affected; the relatively minor differences are caused by the small differences in means of the variables between the eleven years of the study. Of course, this will not be so in all data sets. Our major tool will be the latent covariation matrix, discussed in section 13.3.

The overall stability of the variable domain, $\bar{\phi}$, may be assessed in two different ways. First, Lohmöller gives tables to link the overall stability to the eigenvalues of the first two components from the time mode. With these tables the overall stability is estimated between .85 and .95. This estimate may be compared with the correlations between adjacent occasions (i.e. $r_{k,k-1}$ of the (n×n) correlation matrix R of the time mode). The comparison of the lag-one correlations in Table 13.2 shows good agreement. In addition, it should be observed that the lag-one correlations do not vary much. This leads us to accept the assumption that $\bar{\phi}$ is independent of time, and that by and large the hospitals maintain their overall rank order on the variables over the years.

With a high overall stability the variable components should be very stable as well. Following Lohmöller's guidelines we may infer from $s_{11,11} = 1.71$, and $s_{21,21} = 1.70$ of the latent covariation matrix (Table 13.5A) that both latent variables are equally stable and trait-like. One may seek confirmation for this by inspecting the cross-lag correlations for representative variables (see Table 13.2). Taking the variable *beds* as indicator for size, the stability is obvious; taking the variable *main polyclinical specializations* to indicate degree of specialization, the stability is still clearly visible but rather irregular between year 6 and 8. The cause of the latter is a matter for separate investigation.

The zero value of the covariation between size and degree of specialization for level, $s_{21,11}$, indicates that no cross-lag covariations exist between the latent variables ($\phi_{pp'} = 0$, $p \neq p'$), i.e.Φ is diagonal.

The interaction between level and gain of the two latent variables ($s_{22,11} \neq 0$; $s_{21,12} \neq 0$) shows that the set of variables is not stationary. There is a negative covariation between *level of size* and *gain in degree of specialization* ($s_{22,11} = -2.57$), thus hospitals which are large through the years tend to lose, or at least not gain in degree of specialization, and vice versa. Furthermore, there is a positive covariation between *gain in size* and *level of degree of specialization* ($s_{21,12} = 1.57$) very specialized hospitals tend to become larger, and vice versa.

The importance of the deviations from stationarity are, however, relatively small. This can be assessed from the direction cosines between the combination components as explained in section

Table 13.5 *Hospital study: component covariation matrix*

A. Lohmöller scaling

		size (η_1)		degree of specialization (η_2)	
		level	gain	level	gain
size	level	1.71	0.01	.00	-2.57
	gain	0.01	1.44	1.57	-0.10
degree of specialization	level	.00	1.57	1.70	-0.09
	gain	-2.57	-0.10	-0.09	3.86

B. Standardized

		size (η_1)		degree of specialization (η_2)	
		level	gain	level	gain
size	level	.50	.00	.00	- .02
	gain	.00	.00	.01	.00
degree of specialization	level	.00	.01	.06	.00
	gain	- .02	.00	.00	.00

C. Labelling of elements

		size (1)		degree of specialization (2)	
		level(1)	gain(2)	level(1)	gain(2)
size(1)	level(1)	$s_{11,11}$	$s_{11,12}$	$s_{11,21}$	$s_{11,22}$
	gain(2)	$s_{12,11}$	$s_{12,12}$	$s_{12,21}$	$s_{12,22}$
degree of specialization (2)	level(1)	$s_{21,11}$	$s_{21,12}$	$s_{21,21}$	$s_{21,22}$
	gain(2)	$s_{22,11}$	$s_{22,12}$	$s_{22,21}$	$s_{22,22}$

13.3. From Table 13.5B it follows that the direction cosine between level of size and gain in degree of specialization is -0.02 ($\alpha=91^{\circ}$), and the direction cosine between level of degree of specialization and gain in size is 0.01 ($\alpha=89^{\circ}$). In other words, the deviations from stationarity do not succeed in introducing substantial non-orthogonalities between the combination components.

These small direction cosines may seem strange, when considering the sizes of the elements in Table 13.5A. It should, however, be realized that the Bartussek (1973) scaling of the latent covariation matrix has eleminated the dependency of the elements on the importance of the components they refer to (see also section 6.9). This means that all components are treated on the same level, which was done partly to fit in with the assumption that all latent variables were normalized per occasion. This presentation has some advantage in highlighting the interactions, but at the same time gives a rather incorrect impression of their sizes.

13.6 CONCLUSION

The following (methodological) conclusions can be drawn from the application of three-mode principal component analysis of multivariate longitudinal data.

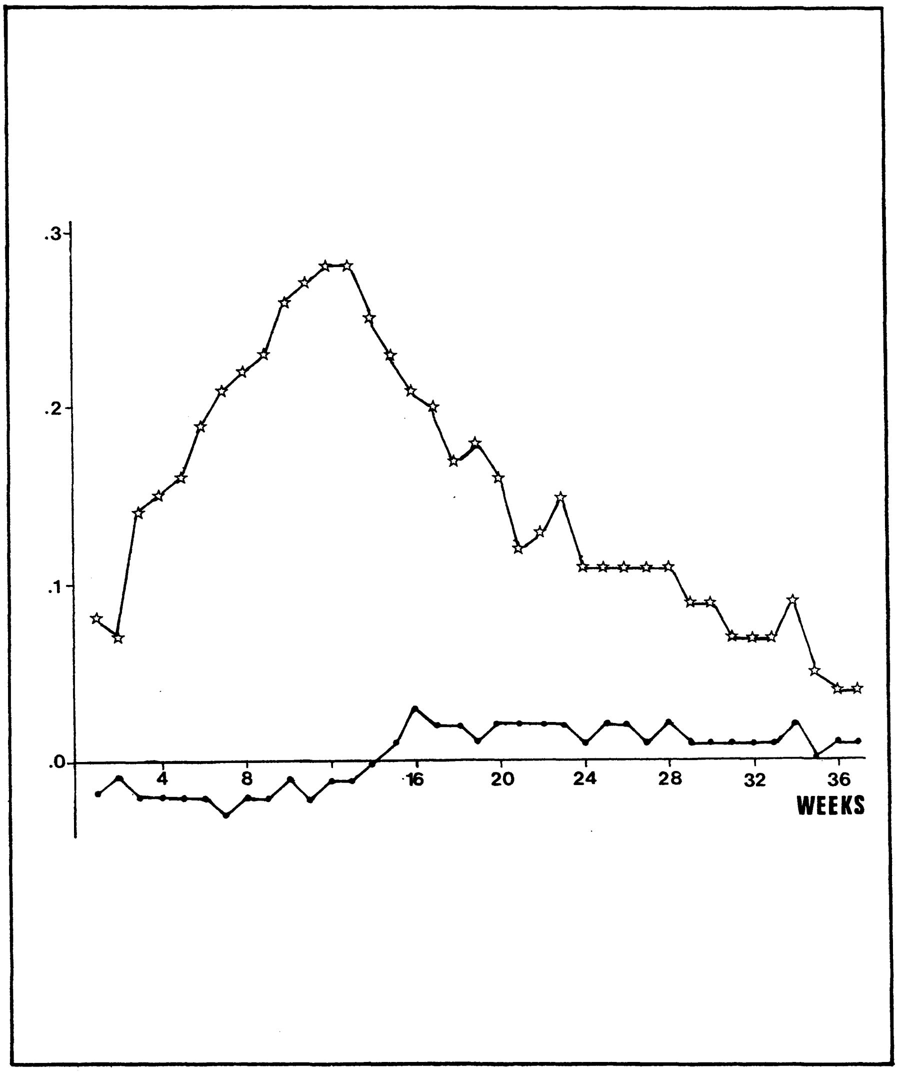
Insight in growth and developmental processes in multivariate longitudinal data can be acquired by inspecting relationships between the components of the three modes: observational units, variables, and occasions. A detailed analysis of the latent covariation matrix can supply information on differential growth patterns if they exist, and the extended core matrix can help to inspect the changes in the interrelationships between the components over time. In fact, the entries in the extended core matrix can be seen as the scores of observational units on combination components of the latent variables and trends. A further level of detail could be introduced by using the extended core matrix to obtain scores on the latent variables for the observational units at each point in time, and by plotting these scores against time, but this suggestion is not pursued here.

A description of data in terms of an autoregressive model has been shown to be acceptable for certain data sets, but first-order processes might be difficult to find. Our example and most of Lohmöller's needed a second-order term. Furthermore, the theory is still very much underdeveloped; only in specific cases detailed statements are possible, and the proper usage of the proposals in three-mode principal components analysis based on alternating least squares methods has not yet been worked out. Nevertheless certain rough results can be obtained which add to the interpretation. At present the most useful aspect is the interpretational framework that an autoregressive model can provide for three-mode analysis of multivariate longitudinal data.

GROWTH CURVES

14

learning-to-read study



14.1 INTRODUCTION

In this chapter we will discuss the three-mode analysis of growth curves. Various proposals to deal with growth curves (reference, learning, or response curves) have been made. Either the exact form of the curve is of interest, or the differences in parameters to describe these curves in different sub-populations which are central to an investigation, or both. For instance, Snee, Acuff, & Gibson (1979) discuss a series of models to deal with univariate growth curves in several populations using Mandel's (1971) proposals. The essence of these proposals, as discussed in section 6.4, is that they consist of an additive main effect model with multiplicative interactions.

The growth curve proper can be analysed by postulating a functional model, and fitting it to the data. After fitting the curves, an analysis-of-variance procedure on the parameters can be used to assess differences between groups, a procedure already proposed by Wishart (1938). For the data to be described logistic regression models (i.e. regression models with a binary counted variable as response variable) with time as independent variable and *slope* and *intercept* as parameters can be fitted to the individual learning curves (Jansen & Bus, 1982); we will relate some of their results to the outcome of the three-mode analysis.

The analysis performed here can best be seen as a multivariate generalization of the growth (learning) curve approach tackled by Tucker (1958, 1966b and Weitzman, 1963). They performed a singular value decomposition (see section 2.2) on the observed raw data. Applications of this approach can, for instance, be found in Kouwer & Hartong (1961), Van Egeren, Headrick, & Hein (1972), McCall, Appelbaum, & Hogarty (1973), Hamel & Netelenbos (1976), Svendsrød & Ursin (1974). Van de Geer (1962) showed how orthogonal polynomials can be used instead of singular value decomposition, and Van Maanen-Feijen (1968) gives a theoretical and empirical comparison between the latter two approaches. That three-mode principal component analysis is a generalization of this approach to learning curves, follows from the observation that the technique is a generalization of singular value decomposition.

14.2 DATA AND PREPROCESSING

In a study to investigate the process of learning to read seven first-grade children were tested weekly (except for holidays) with five different tests (see Table 14.1 for a description), which were designed to measure different aspects of reading ability. We will not discuss the theoretical rationale behind the test the details of the design, the testing procedures, and the overall quality of the data (see Bus, 1982, Jansen & Bus, 1982, and Bus & Kroonenberg, 1982). Of the seven children which took the tests, one is not included in the present analysis, as he was added to the study at a later moment, and accounted for a large part of the missing data.

Table 14.1 *Learning to read study: Description of tests*

Test	Description
P	regular orthographic short words
Q	regular orthographic long words
R	irregular orthographic long and short words
S	regular orthographic long and short words within context
L	letter knowledge test

Because the tests had different ranges, either 10, 15 or 47 items, the data were rescaled so that all the tests ranged from 0 to 1. In this way all the differences in variation were maintained in the data, while making the tests comparable. Subsequently we constructed the average learning curve by averaging over pupils and tests for each occasion. Thus, in effect, we will use a mixed additive and multiplicative model for the data (see section 6.4 for a discussion of such models) analysing the common part of the learning curve additively, and the residuals, ε_{ijk}, multiplicatively:

$$z_{ijk} = \mu + \gamma_k + \varepsilon_{ijk} \qquad (i=1,\dots,\ell;j=1,\dots m;k=1,\dots n)$$

$$= \mu + \gamma_k + \sum_{p=1}^{s} \sum_{q=1}^{t} \sum_{r=1}^{u} g_{ip}h_{jq}e_{kr}c_{pqr},$$

with μ the overall average, and γ the occasion main effect with the restriction $\Sigma\gamma_k = 0$. The least squares estimates for $\mu = \bar{z}_{...}$, and for $\gamma_k = \bar{z}_{..k} - \bar{z}_{...}$ $(k=1,\dots,n)$ according to the standard theory of linear models. Thus the residuals $\varepsilon_{ijk} = z_{ijk} - \bar{z}_{..k}$ $(i=1,\dots,\ell; j=1,\dots,m;k=1,\dots,n)$.

An advantage of the above model is that the (cor)relations between pupils and tests are no longer influenced by the average growth curve, and the interactions between tests and pupils over time can be analysed separately. Of course, averaging and interpreting the average growth curve is only meaningful if the individual curves more or less resemble one another (see e.g. Tucker, 1966b, p.480-483, and the references therein for a discussion of problems around average learning curves). Furthermore, averaging over tests is only meaningful if all tests measure essentially the same variable to a different extent. In the present case this is not unreasonable considering the high intercorrelations (average = 0. 88; taken over all time-pupil combinations).

A study by Jansen & Bus (1982) on the same data suggests that per test a logistic regression model with the same slope for all individuals is not an unreasonable model. The slopes are, however, different across tests. This implies that the average learning curve, $\mu + \gamma_k$ $(k=1,\dots,n)$ does not necessarily represent any one test in particular. Nevertheless, it serves as a baseline for comparisons between tests and pupils.

Removing only the time main effect has the advantage that the difference in complexity between tests, and differences in reading ability between pupils remain in the analysis of the interactions. One could, of course, remove the pupil and test main effects as well, and analyse the remaining residual with three-mode principal component analysis, but this course is not pursued here.

14.3 AVERAGE LEARNING CURVE

In Fig. 14.1 the $\bar{z}_{..k}$ $(k=1,...,n)$ have been plotted against k (time) to give a general impression of the shape of the learning curves. There is a rapid increase in the test scores until about week 16, and a gradual growth until the ceiling of one, i.e. all items of the tests correct. The actual process of learning-to-read thus roughly takes place in roughly those first 16 weeks; the later weeks are used to perfect the reading of the better pupils, and remedy that of the weaker ones.

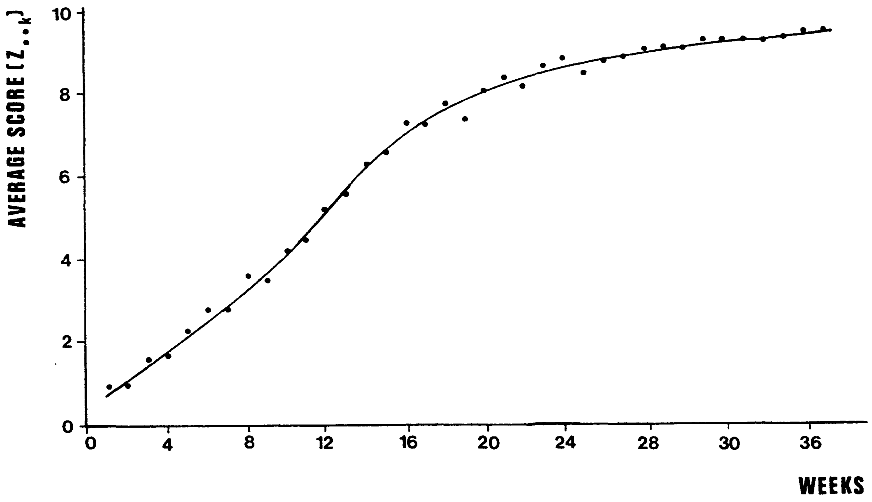


Fig. 14.1 *Learning-to-read study: average learning curve*

14.4 GENERAL CHARACTERISTICS OF THE SOLUTION

The primary research questions in this study were centred around questions like: Was the performance of the pupils uniform over tests, did the tests cause uniform differentiation between pupils, or were certain pupils better on some tests, while others were better on other tests?

To answer these questions we will have to look at joint plots (see section 2.4, and 6.10) of the pupils and the tests. Before we can do this it is necessary to investigate how many dimensions are necessary for each of the three modes. At least two are needed for the pupils and the tests each (see Table 14.2). The table also shows that for occasions one component is sufficient, especially if one remembers that the most important source of variation (i.e. due to the average learning curve) has already been removed. Thus the first time component already explains about 70% of the interactions. Fig. 14.2 shows the curves of the first and second components (scaled according to their relative weight). The second component seems to exhibit little (interesting) variation, and we will, therefore, not discuss it further.

Table 14.2 *Learning-to-read study: proportions explained variation*

mode	proportion explained variation of components 1	2
1 pupils	.44	.33
2 tests	.43	.34
3 occasions	.70	.07

The first component has a nearly perfect rank correlation with the variation per occasion measured by the fitted sum of squares. In other words, high loadings correspond to large differences between tests and/or pupils (the individual curves show that both is the case), and small loadings to small differences. From Fig. 14.2 we may conclude that in the first 12 weeks large differential growth exists between pupils on the tests and that the differences

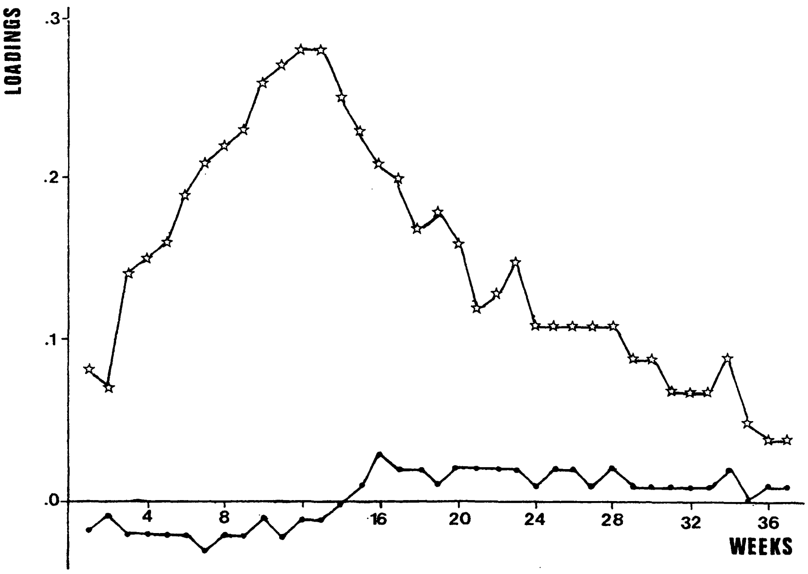


Fig. 14.2 *Learning-to-read study: Trends*
(✫ = first trend; • = second trend)

have largely disappeared at the end of the schoolyear. Inspection of the individual curves and observations in the classroom show that the decreasing differences are largely a ceiling effect, in other words, that further progress was impossible to measure with these tests as the better (quicker) pupils already obtained maximum scores on one or more tests.

14.5 ANALYSIS OF INTERACTIONS

With respect to the interactions between the tests and pupils, the foregoing shows that we only need to look at their interrelationships in the first core plane of a TUCKALS3 analysis (see Table 14.3), which shows the combinations of the components of tests and pupils responsible for the differential growth curves. The severe non-diagonality of this plane suggests that the relationships are rather complex. However, the near equality of the absolute values of the off-diagonal elements shows that the two sets of axes (of tests, and of pupils) are rotated with respect to one another over an angle of approximately 45° (see note of Table 14.3 and section 6.9).

Table 14.3 *Learning-to-read study: relationships between test and pupil components*

first frontal plane of Tuckers3 core matrix

		test components		proportion explained variation	
pupil	1	17.3	-11.4	.27	.12
components	2	11.2	14.7	.12	.70
				sum	.70

Note: direction cosine between pupil and test components
$\cos \alpha = 11.2 / \sqrt{17.3 \times 14.7} = .70 \rightarrow \alpha \cong 45^\circ$.

In the joint plot corresponding with the first TUCKALS3 core plane (Fig. 14.3) the approximate positions of the original axes are drawn. How may we interpret Fig. 14.3 as far as the interactions of tests and pupils is concerned, or how do the combinations of components contribute to the differential growth? The interpretation is simplest when we take the component axes of the tests as reference. On the first axis we find a large differentiation between the tests, but not between pupils. In other words, the first test axis shows that all pupils have a positive differential growth (i.e. have curves above the average learning curve, see Fig. 14.1) for tests with positive loadings on this component (S,P,L), and a negative differential growth (i.e. curves below the average learning curve) for tests which have negative loadings there (R,Q). Thus, for all pupils with only minor deviations, S,P,L,Q,R are ordered from simple to difficult.

On the second test axis the situation is exactly reversed. There the tests have about equal loadings, but there are large differences between pupils. Pupil 4 is best on all tests followed

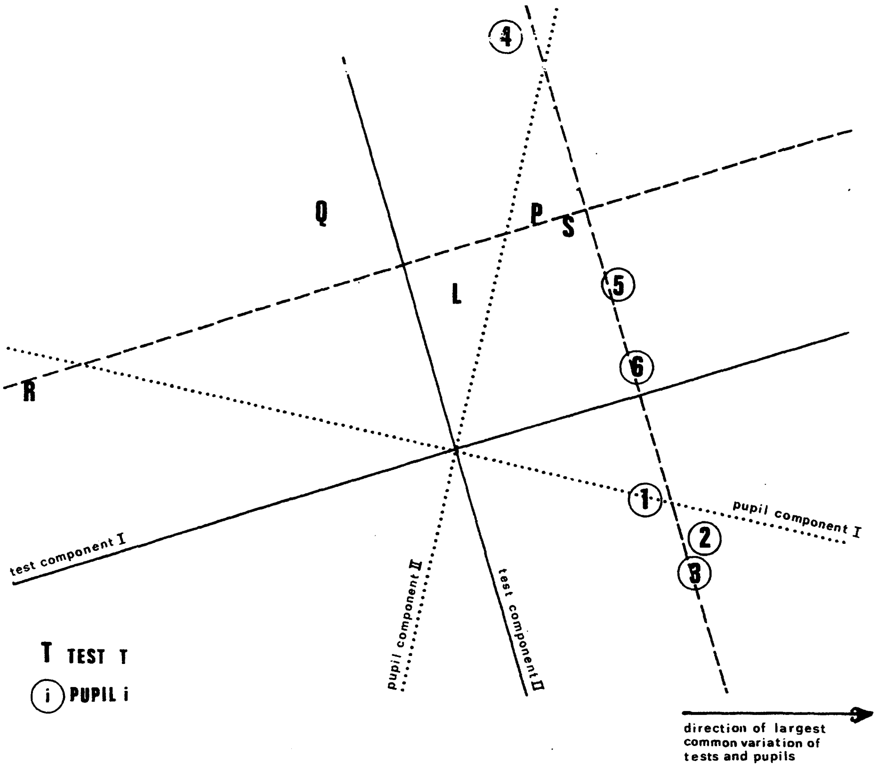


Fig. 14.3 *Learning-to-read study: Joint plot for tests and pupils*

by 5, 6, 1, 2 & 3. We see that 4 and 5 show positive differential growth on virtually all tests, while 6 lies more or less on the average curve for all tests, and 1, 2, and 3 show a negative differential growth on all tests. The slight variation in the loadings of the tests on this axis indicates that the spacing of the pupils is not exactly the same on all tests, but only approximately so. In particular, Q and L are somewhat different from the other tests.

In summary, there seems to be more or less independence of the differential growth due to differences in tests, and the differential growth due to differences in pupils. In other words, a test is easy or difficult for all pupils, and a pupil is better or worse than another pupil on all tests. This independence is nicely confirmed by Table 4 of Jansen & Bus (1982), which shows the estimated points on the time axis, $t_{ij}^{(\frac{1}{2})}$, at which a performance level of .50 (for the range 0-1) is reached. The estimates are based on logistic regression curves fitted to the scores of each pupil on each test (except for L, which was not included in their study). In Table 14.4 we have reproduced her table in rearranged form, as well as the residuals $r_{ij}^{(\frac{1}{2})}$ after we have fitted a two-way main effect model to these half-way scores $t_{ij}^{(\frac{1}{2})}$:

Table 14.4 *Learning-to-read study: estimated half-way scores* (in weeks)

							residuals from two-way main effects model			
	tests						tests			
pupils	S	P	Q	R	mean	row effect	S	P	Q	R
4	4	4	2	15	6	-11	3	2	-2	-1
5	9	11	12	24	14	- 3	6	1	0	0
6	12	12	19	27	18	+ 1	-1	-2	3	-1
1	14	15	19	28	19	+ 2	0	1	2	-1
3	14	17	24	33	22	+ 5	-3	-1	4	1
2	18	17	23	34	23	+ 6	0	-2	2	1
mean	12	13	15	27	17					
column effect	-5	-4	-2	+10						

$$r_{ij}^{(\frac{1}{2})} = t_{ij}^{(\frac{1}{2})} - t_{..}^{(\frac{1}{2})} - (t_{.j}^{(\frac{1}{2})} - t_{..}^{(\frac{1}{2})}) - (t_{i.}^{(\frac{1}{2})} - t_{..}^{(\frac{1}{2})}),$$

with $t_{ij}^{(\frac{1}{2})}$ the time at which pupil i scores .50 on test j; the "." indicates averaging over the index it replaces. The residuals do not seem to behave in a systematic way except that Q has maybe too many positive scores, confirming the slightly different behaviour of Q which we noted above. From the lack of pattern in the residuals we may therefore conclude that no interactions exist between pupils and tests with respect to the half-way scores.

14.6 CONCLUSION

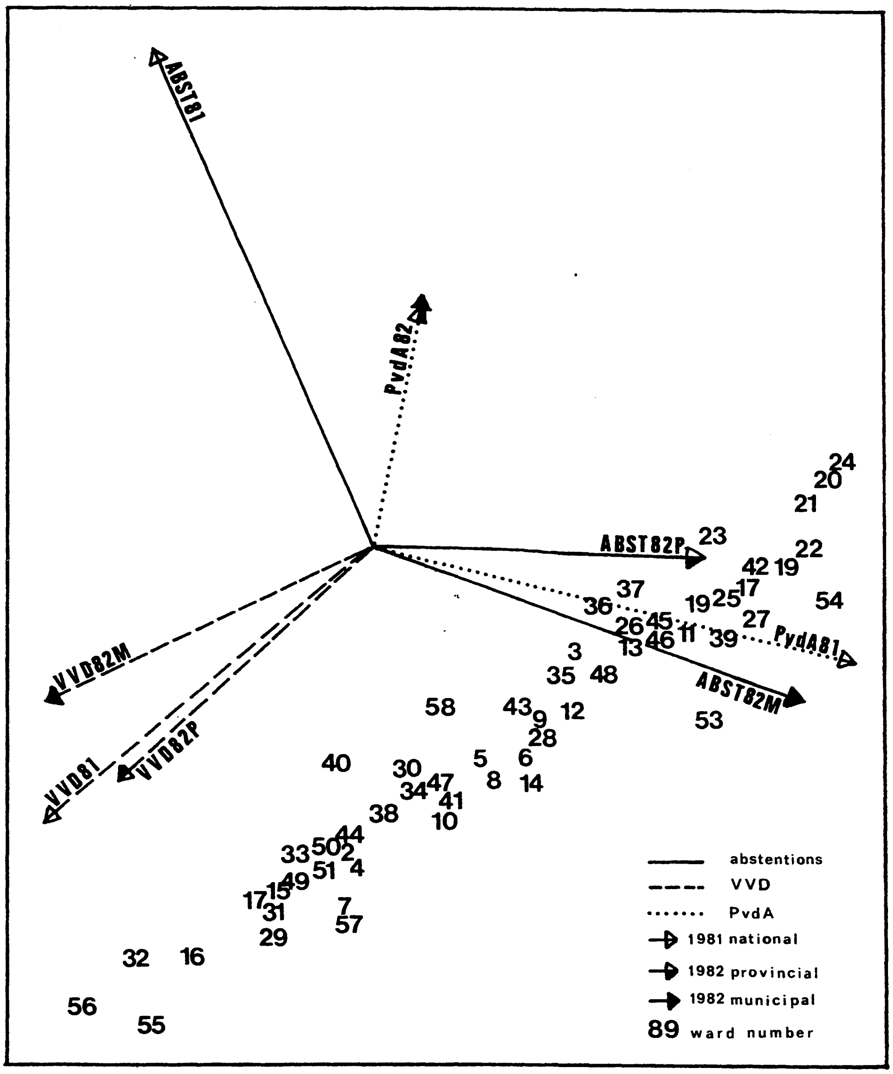
In conclusion we may say that the difference in the scores of the pupils and the tests can be ascribed to two more or less independent factors: the varying degrees of difficulty of the tests, and the differences in ability of the pupils. Furthermore, the relationships between tests and pupils are time-independent, and the time aspect is contained in the loadings of the first time component. In other words, the structural relationships are invariant over time.

On a methodological level the example shows how three-mode principal component analysis can be used to analyse learning curves. It also gives a demonstration of the use of mixed additive and multiplicative models. Finally, the individual curve fitting as performed by Jansen & Bus (1982) and the three-mode analysis presented here, as well as a similar three-mode analysis by Bus & Kroonenberg (1982) nicely supplement each other. A more precise statement about the ways the two techniques can be used in conjunction requires another study.

THREE-MODE CORRESPONDENCE ANALYSIS

15

Leiden electorate study



15.1 INTRODUCTION

Contingency tables turn up in many research projects in many contexts, and there exists an extensive collection of techniques for their analysis. Especially in recent years the development of loglinear models for contingency table analysis has enabled researchers to make more detailed statements about association in multi-way tables than just reporting descriptive levels of significance or p-levels. Notwithstanding, or because of, the refined machinery connected with loglinear models there are serious problems with their application to large tables, and/or to higher dimensional tables. These problems centre around (1) the null distribution and power of test statistic when the numbers of observations per cell is low, (2) the difficulty of interpreting the interaction terms when there are very many of them (as is the case with large tables), and (3) the complexity of interpreting high-order interaction terms, especially if there are a lot of observations.

Here we will specifically pay attention to interactions of large three-way contingency tables.

15.2 LOGLINEAR MODELS, INTERACTIONS, AND CHI-TERMS

A saturated model for two-way contingency tables is a model which completely accounts for the data by specifying all effects and interactions. It has the form

$$\log f_{ij} = \ell_{++} + (\ell_{i+} - \ell_{++}) + (\ell_{+j} - \ell_{++}) + \log r_{ij}, \quad (i=1,\ldots,\ell;j=1,\ldots,m)$$

with f_{ij} the observed cell count; ℓ_{ij} = log f_{ij}; r_{ij} is the residual and the "+" indicates summation over the index it replaces. Using Bishop, Fienberg & Holland's (1975) notation, this can be written as:

$$\log f_{ij} = u + u_{1(i)} + u_{2(j)} + u_{12(ij)} + \log r_{ij} \ .$$

There are two main effect vectors u_1 and u_2, and one two-way interaction matrix u_{12}. The formula for a non-saturated model has a combination of one or more of these terms on the right hand side. The most common model for a two-way table is the model of independence between rows and columns:

$$\log e_{ij} = u + u_{1(i)} + u_{2(j)}.$$

This model may be tested against the data by assessing the size of the residuals $r_{ij} = f_{ij}-e_{ij}$ via Pearson 's X^2-test,

$$X^2 = \Sigma\ (f_{ij}-e_{ij})^2/e_{ij},$$

or the -2loglikelihoodratio,

$$LR= -2f_{ij}\log(f_{ij}/e_{ij}).$$

The values of the test statistics are evaluated against percentage points of the χ^2-distribution with $(\ell-1)(m-1)$ degrees of freedom. Given non-independence, one can inspect the residuals for specific patterns. While these patterns are easily visible in small tables, visually analysing more or less subtle relationships from a large table can become too difficult. In addition, the residuals themselves suffer from differences in size due to original differences in size of the frequencies, and for comparing the residuals it is more appropriate to standardize them in some way. One obvious way is to use *standardized residuals* or as we, somewhat inappropriately, will call them *chi-terms*:

$$X_{ij} = (f_{ij}-e_{ij})/e_{ij}^{\frac{1}{2}}.$$

A more subtle kind of standardization leads to Haberman's *adjusted residuals* (Haberman, 1976). Here we will deal exclusively with the chi-terms, or X_{ij}.

When confronted with a two-way contingency table a reasonable procedure for analysis can be summarized as follows:

a) construct a model;
b) test for appropriateness of the model (optional);
c) interpret the terms of the model;
d) compute the chi-terms;
e) analyse the chi-terms for specific patterns.

For three-way tables the procedure is essentially the same, but the situation is more complex as there are now three main effects, three two-way interactions and one three-way interaction:

$$\log f_{ijk} = u+u_{1(i)}+u_{2(j)}+u_{3(k)}+u_{12(ij)}+u_{13(ik)}+u_{23(jk)}+u_{123(ijk)}.$$

Again a model consists of a subset of terms from the right hand side. A simple model is the *three-way independence model* consisting of u, u_1, u_2, and u_3. In this case we obtain the chi-terms:

$$X_{ij} = (f_{ijk}-e_{ijk})/e^{\frac{1}{2}}_{ijk}, \text{ with}$$

$$\log e_{ijk} = u+u_{1(i)}+u_{2(j)}+u_{3(k)},$$

$$\text{or} \quad e_{ijk} = f_{i++}f_{+j+}f_{++k}/ f^2_{+++}$$

Looking at these chi-terms or standardized residuals implies inspecting all two-way and three-way interactions simultaneously. One can also use the chi-terms from other loglinear models which include more u-terms, and we will do so in our example.

15.3 CORRESPONDENCE ANALYSIS FOR CONTINGENCY TABLES

In this section we will first give a short summary of correspondence analysis of two-way contingency tables, and then show how it can be generalized to three-way tables. Further extensions to higher-way tables are possible but will not be discussed here.

Two-way tables. The aim of correspondence analysis (Benzécri, 1976; Gifi, 1981, Ch. 3) for two-way contingency tables is to quantify or scale the column and row categories in such a way that the nature of the interaction both within and between rows and columns becomes directly visible, for example via a combined plot of the row and column categories. The way to arrive at such a combined plot is via the singular value decomposition $G \Lambda H'$, (see section 2.2), of the matrix X with chi-terms $(X_{ij} = (f_{ij}-e_{ij})/e_{ij}^{\frac{1}{2}})$. This procedure has close links with the joint plots considered in section 6.10. The columns of G are the quantifications (or scalings) of the rows, and columns of H are the quantifications of the columns of the table. The combined configuration may be interpreted as follows (see e.g. Israëls, et al., 1981):

a. the *origin* represents the marginal distributions for both rows and columns, and thus has the same meaning for both, i.e. it is the centre of gravity for each row (column);

b. the *distance* of a row (column) point to the origin shows to what extent the distribution within that row (column) deviates from the marginal distribution. Thus the origin is the point indicating independence between the row and column variables;

c. the *direction* emanating from the origin indicates the kind of deviation; when row (column) categories are close together their conditional distributions resemble each other;

d. when rows and columns deviate in opposite *directions* they are negatively related;

e. the *size of the deviation from independence* is indicated by the distance from the origin.

Central to these interpretations are the ideas of distance, direction, and the role of the centre of the configuration. For instance, Gifi (1981, p. 134-137) gives precise definitions of the distances both in terms of deviations from the marginal proportions, and in terms of "χ^2-distances" between rows (columns) of the contingency table. The role of the origin, as the point representing the marginal distributions and as the point indicating independence, can be seen directly from the fact that the analysis is performed on the matrix of chi-terms rather than on the original frequencies.

Three-way tables. The generalization to three-way tables is based on the observation in section 2.2 that three-mode principal component analysis is a generalization of singular value decomposition. A three-mode principal component analysis on a matrix of chi-terms derived from some three-way contingency table is, therefore, a three-way analogon of correspondence analysis, as was pointed out by De Leeuw (1981, pers. comm.). Another kind of application of multi-mode models to contingency tables can be found in Carroll (1975), Green, Carmone, & Wachspress (1976), Carroll, Pruzansky, & Green (1977). They used the CANDECOMP procedure (Carroll & Chang, 1970; see also section 3.3) to estimate the parameters of Lazarsfeld's latent class model (Lazarsfeld & Henry, 1968). In Carroll et al. (1977) it was shown that their procedure is also a generalization of correspondence analysis.

It is possible to make another fruitful generalization within the context of correspondence analysis to three-way tables. One may analyse a table of chi-terms not only from an independence model, but from any model. Deviations from the origin are then interpreted as deviations from this model.

It is not the intention to develop the mathematics of this proposal. Instead, we will show via an example of the voting behaviour in Leiden (the Netherlands) during three elections how the procedure works.

15.4 LEIDEN ELECTORATE STUDY: DATA

In the Netherlands there are three kinds of elections: national, provincial and municipal, which follow each other mostly at irregular intervals. Although different issues play different roles in these elections, each has a national overtone. Popularity and impopularity of political parties at the national level invariably have their impact on the local elections, and on the other hand, results from local elections are seen as a measure for the popularity of parties on the national level.

The data for the present study consist of the results for the wards of Leiden in these three different kinds of elections: the 1981 national parliamentary elections, the March 1982 provincial elections, and the June 1982 municipal elections. Data are available for the *58* wards or precincts of the city and *9* parties or combinations of parties who participated in all *3* elections. Van der Heijden (1982) analysed the first two elections (with *10* rather then *9* parties) using correspondence analysis as implemented in ANACOR (Gifi, 1981 (Ch.3), 1982). In particular, he performed regular correspondence analysis on the three-way table by rearranging the table as an (58+58)×10 table, a 58×(10+10) table, and an (incomplete) (58+10+2)×(58+10+2) bi-marginal table. See also Gifi (1981, section 4.4.4) for similar correspondence analyses on three-way tables. Here we will analyse the 58×9×2 table in its three-way form. The three zero cells of the 1982 municipal elections were set equal to one to facilitate comparison with other analyses.

15.5 LOGLINEAR ANALYSIS

Before embarking on a three-mode correspondence analysis, it is useful to investigate first a three-way contingency table with loglinear models. This kind of analysis gives insight in the relevance of the various interactions present in the data, and makes it possible to decide which interactions should be investigated further.

Only those models are appropriate or permissible for the Leiden electorate data which include the ward, election, and ward × election effects, as their marginal distributions are fixed by the design (see e.g. Fienberg, 1980, p.95,96 for a discussion of margins fixed by the design), as the size of the wards and the total number of eligible voters in each election are known a priori. As, in addition, the sizes of the wards in each election were almost the same, the variables *ward* and *election* are independent as well. According to Theorem 2.4-1 of Bishop, Fienberg & Holland (1975, p.39; see also Fienberg, 1980, p.49) this implies (together with the negligeable three-way interaction; see below) that we may collapse either over wards to investigate the party×election interaction, or over elections to investigate the party×ward interaction. In other words, inspecting and interpreting these two-dimensional margins is all that is necessary for these data, possibly with a quick look at the three-way interaction to make sure everything is in order there.

All possible models were generated by BMDP4F (Dixon, 1981, p.143ff), and for each the -2loglikelihood ratio statistic with the appropriate degrees of freedom was determined. In Fig. 15.1 we have summarized the results of all possible hierarchical models, and in Table 15.1 we have given some details for the permissible models.

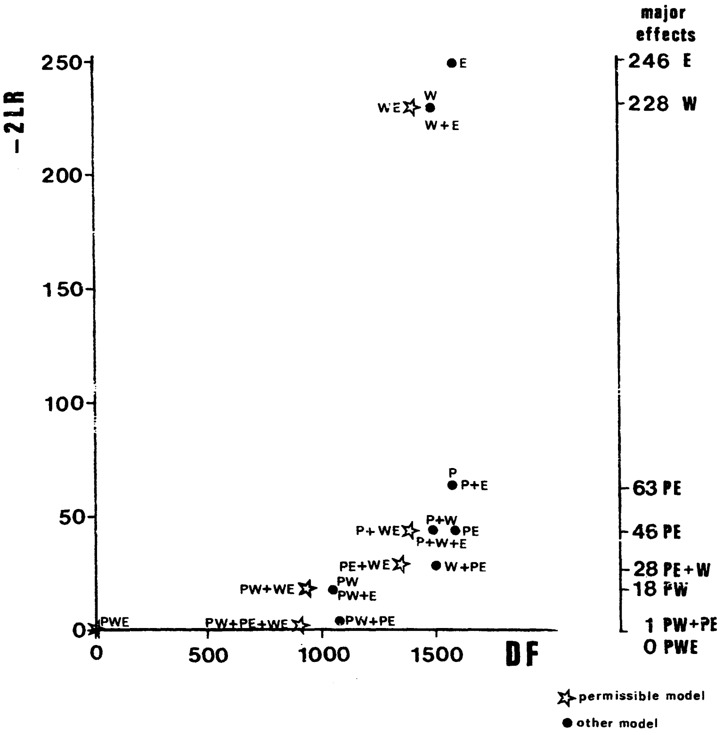


Fig. 15.1 *Leiden electorate study: Fit of loglinear model*

The huge values of the test statistic ensure the significance of each and every model, i.e. no model fits the data well. With a total of roughly 232 thousand observations and 1566 cells this is hardly surprising. The large number of observations even ensures that uninteresting small differences between models lead to significant test results, especially between the levels indicated in

Table 15.1 *Leiden electorate study: permissible loglinear models*

models	degrees of freedom	-2 loglikelihood ratio (in thousands)	
W+E	1480	228	
WE	1368	228	
P+WE	1360	45	Model I (three-way independence)
PE+WE	1344	28	Model II (PW+PWE interactions)
PW+WE	912	19	Model II (PE+PWE interactions)
PW+PE+WE	896	1	(PWE interaction)
PWE	0	0	saturated model

Note: The models are designated by the highest order interactions present. Implicit in the notation is the inclusion of the hierarchical lower-order interactions, e.g. WE indicates the model W + E + WE.

Fig.15.1. In other words testing is an uninteresting, but especially unnecessary exercise in this case. What is interesting, is the relative magnitudes of the -2loglikelihood ratios themselves.

Fig. 15.1 shows that the major effects are *party, party×ward*, and *party×election*. It also shows that the three-way interaction can be neglected, i.e. set to zero, for most practical purposes. In other words, we may here assume that the three-way interaction reflects random variation. After all, it takes 896 degrees of freedom to reduce the test statistic from 1264 to zero, when going from the model including all two-way interactions to the model which also includes the three-way interactions, i.e. the saturated model.

Inspecting the ward×party (58×9) two-dimensional margin is clearly something that cannot be done properly without an adequate graphical represention. In fact an ordinary two-way correspondence analysis on the party×ward margin is already sufficient (see Van der Heijden, 1982). A visual inspection of a table of the party× election interaction is, on the other hand, quite feasible as Table 15.2 shows. The most salient features are the decline of the PvdA (and the CDA and D'66 to a lesser extent), the relative stability of the VVD, and the other smaller parties, and the 150%

Table 15.2 *Leiden electorate study: Model III: party×election interaction*

party	Average number of votes per party over all wards			standardized residuals (obs-exp)/sqrt(exp)		
	National 1981	Provincial 1982	Municipal 1982	N 1981	P 1982	M 1982
PvdA	404	245	250	6.3	- 3.3	- 2.9
CDA	221	180	178	2.2	- 1.0	- 1.1
VVD	211	207	193	0.7	0.1	- 0.8
Abstentions	187	468	506	-10.0	4.0	5.9
D'66	146	93	55	5.0	- 0.6	- 4.4
Small left	114	113	111	0.3	- 0.0	- 0.2
Small right	20	21	18	0.1	0.3	0.4
Invalid	9	9	6	0.4	0.3	- 0.7
Soc. Party	8	10	26	- 1.7	- 1.2	2.9

Note: small left: CPN, PSP; small right: BP, GPV, SGP

increase in abstentions. After the provincial and municipal elections the abstainers were by far the largest party, and the S(ocialist) P(arty) also increased its support considerably. Furthermore, in absence of a sizeable three-way interaction, the conclusion may be drawn that the abstainers came from especially the PvdA wards. The great decline of the PvdA in the provincial, and the slight increase in the municipal elections follow the popularity of this party on the national level quite closely.

For the purpose of illustrating three-mode correspondence analysis, and in order to compare its results with the loglinear results we will investigate the chi-terms (or standardized residuals) of a number of models via this technique. This does not imply that for these data all the models should be investigated in this way, if it was the subject matter which was of primary importance. We will, however, investigate the chi-terms from the models I, II, and III in Tabel 15.1.

Determining an appropriate model for the data, and inspecting the chi-terms or standardized residuals from such a model is, by the way, directly analoguous to the problem of determining the appropriate centring for numerical data treated in detail in Chapter 6.

15.6 MODEL III: POLITICAL ALIGNMENT OF THE CITY OF LEIDEN

In this section we will look especially at the party×ward interaction. In other words, we want to find out how people in the various wards voted. This can be accomplished by looking at the party×ward margin, i.e. summed over elections (external averaging), or by using Model III:

$$\log e_{ijk} = u + u_{W(i)} + u_{P(j)} + u_{E(k)} + u_{WE(ik)} + u_{PE(jk)} \text{ , or}$$

$$e_{ijk} = f_{i+k}\, f_{+jk}\ / f_{++k} \text{ ,}$$

with W=ward, P=party, and E=election. In this case only the party× ward interaction, $u_{WP(ij)}$, (i.e. which parties acquire their votes from which wards) and the three-way interaction $u_{WPE(ijk)}$ remain to be analysed via a correspondence analysis of the chi-terms. By looking at the joint plot from a Tucker2 analysis using the average

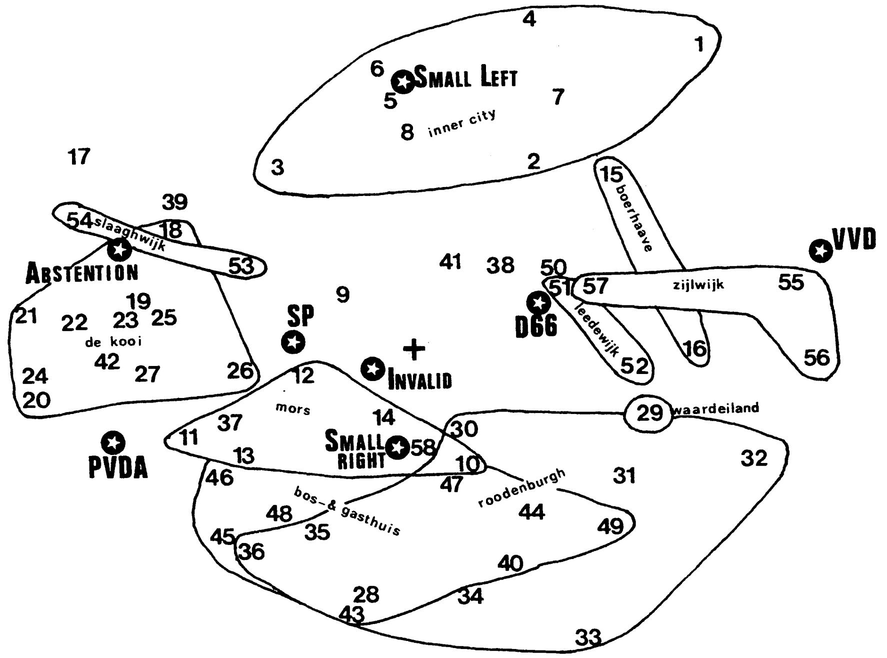


Fig. 15.2 *Leiden electorate study: Model III - party×ward interactions* (labelled by ward number)

core plane (internal averaging) we can display the relationships between the wards and the parties (Fig. 15.2). This plot can be interpreted using the guidelines given in section 15.3.

Invalid votes occur apparently randomly over the city, and are therefore located near the zero point of the plot. Wards 55 and 56 both vote predominantly VVD, the right-wing conservative party, more than the Leiden average. The inner city (wards 1 through 8) votes more than other wards for the small left-wing parties, but, in addition, the VVD receives an excess of votes in ward 1, the PvdA in ward 3. Similarly, in a typical labour district such as 'De Kooi' (wards 18 to 26) the labour party PvdA receives more votes than its marginal distribution would predict. A number of interesting details can be discerned, but these are probably only meaningful for people familiar with Leiden itself, and they will not be covered here.

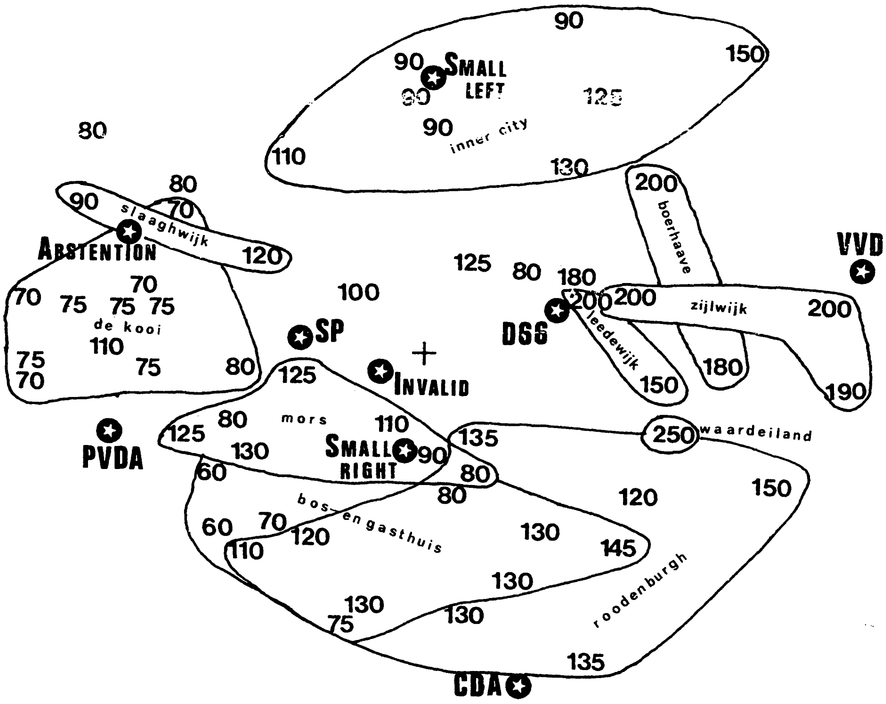


Fig. 15.3 *Leiden electorate study: Model III - real estate values and party alliances* (in thousands of guilders; aug. 1982)

In Fig. 15.3 the wards are labelled according to the estimated average value of the real estate in these wards. From this plot it is clear that average real estate value is a reasonable global predictor of voting behaviour, but it should be kept in mind that in some wards the variance of real estate values is very large.

So far we have concentrated on the party×ward interaction, and have disregarded the three-way interaction on the assumption that it was not overly large. From the TUCKALS2 core matrix (Table 15.3) we are able to assess, whether this is indeed a valid assumption. In all elections the ward×party relationships turn out to be virtually the same, as do the relative SS(Fit) of each election, and thus the lack of a substantial three-way interaction is confirmed.

Table 15.3 *Leiden electorate study: T2 core matrix for Model III*

		election							
		national (1981)		*provincial (1982)*		*municipal (1982)*		*average*	
components for:				parties				parties	
		1	2	1	2	1	2	1	2
wards	1	-18	0	-17	1	-18	- 1	-18	0
	2	0	11	- 1	10	1	10	- 0	10
Relative SS (Fit)		.78		.79		.79		Overall=.79	

15.7 MODEL II: DECLINE OF SUPPORT FOR THE PvdA

As mentioned before, the party×election interaction can easily be inspected from the two-dimensional margin as we did in Table 15.2. For illustrative purposes, however, we very briefly report the analysis of chi-terms from Model II:

$$\log e_{ijk} = u + u_{W(i)} + u_{P(j)} + u_{E(k)} + u_{WP(ij)} + u_{WE(ik)}, \text{ or}$$

$$e_{ijk} = f_{ij+}\, f_{i+k} \,/ f_{i++}$$

with again W=ward, P=party, and E=election. The chi-terms contain the two-way interaction between party and election, and the three-way interaction, as they are the only interactions not included in the model. In a TUCKALS2 analysis, the unimportance of the three-way interaction now should follow in a TUCKALS2 analysis from the unidimensionality of the ward space, with all entries approximately equal to $1/\sqrt{58} = 0.13$. The real average of the first component was also 0.13 with a standard deviation of 0.03. All wards were within ± 2 standard deviations from the average. The overall SS(Fit) was 0.940, and that of the first ward component 0.933, confirming the uni-dimensionality of the ward space.

Before reporting the results for the party×election interaction we will show the consequence of the uni-dimensionality of the ward space (G), and the equality of all g_i:

$$\hat{z}_{ijk} = \sum_{p=1}^{s} \sum_{q=1}^{t} g_{ip} h_{jq} c_{pqk} = \sum_{q=1}^{t} \bar{g}_1 h_{jq} c_{1qk} \propto \sum_{q=1}^{t} h_{jq}\, c_{1qk},$$

in other words, for each ward i we obtain the same values for the party×election interaction. Adjusted with a constant scaling factor these values are given in Table 15.4.

Table 15.4 *Leiden electorate study: Model II - party×election interactions*

party	National 1981	Provincial 1982	Municipal 1982
PvdA	6.3	-3.4	-2.9
CDA	2.2	-0.9	-1.3
VVD	0.7	0.2	-0.8
Abstentions	-9.9	3.9	6.0
D'66	5.1	-0.7	-4.4
Small left	0.3	0.0	-0.3
Small right	0.1	0.4	-0.5
Invalid	0.2	0.4	-0.6
Soc. Party	-1.7	-1.2	2.8

Note: The values supplied by the program were divided by 2.3.

Comparison with the standardized residuals in Table 15.2 shows that the two are identical.

15.8 MODEL I: SIMULTANEOUS ANALYSIS OF ALL NON-FIXED INTERACTIONS

Model I is the independence model, i.e. it postulates that the expected values of the cell counts can be reconstructed from the one-dimensional margins, and all fixed margins. This results in independence between party and ward×election. Model I has the form:

$$\log e_{ijk} = u + u_{W(i)} + u_{P(j)} + u_{E(k)} + u_{WE(ik)}, \text{ or}$$

$$e_{ijk} = f_{i+k} f_{+j+} / f_{+++}$$

with W=ward, P=party, and E=election. The chi-terms for this model consist of the ward×party and the party×election, and the three-way interactions. If we neglect the three-way interaction, the correspondence analysis of the chi-terms allows us a simultaneous analysis of the most important interactions in the data.

Let us first inspect the party space (Fig. 15.4) independently of the wards and the elections, as a proper interpretation of its structure is crucial for the remainder. The space is dominated by the position of the VVD, the PvdA, and the abstentions. In other words these are the 'parties' with conditional distributions substantially deviating from those expected under Model I. From the previous analyses we know the party space to be a resultant of the party×ward and the party×election interactions. However, from this party space alone we may conclude that the VVD, PvdA, and the abstentions are all negatively related and in the same way, as the angles between their vectors emanating from the origin are all roughly equal, i.e. 120°. In Fig. 15.4 the first component of the party space for Model II and Model III are drawn on the basis of a procustes rotation (see e.g. Gower, 1975).

Table 15.5 *Leiden electorate study: TUCKALS2 core planes*

elections

	national (1981)		provincial (1981)		municipal (1982)	
components for:			parties			
	1	2	1	2	1	2
wards 1	- 1	-16	-13	- 1	-16	- 1
2	-17	3	- 4	-18	1	- 9

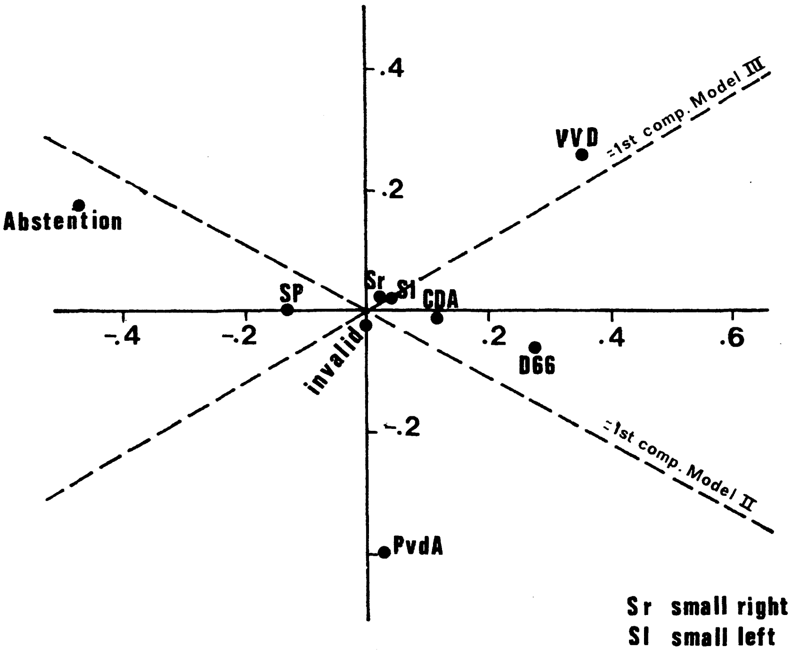


Fig. 15.4 *Leiden electorate study: Model I - party space*

The angles between the first components of Models I, II, and III (computed from separate procrustes rotations) are $\alpha_{I,II} = 29^{o}6$; $\alpha_{I,III} = 31^{o}8$; $\alpha_{II,III} = 49^{o}9$. The correlations between the first components after the separate rotations are $r_{I,II} = .77$; $r_{I,III} = .82$; $r_{II,III} = .45$.

The TUCKALS2 core matrix (Table 15.5) shows how for each election the configurations of parties and wards are connected. Note that the minus signs indicate that the components of the wards and parties are inversely related.

If we neglect the two smaller values in each frontal plane, we can see the dramatic change which took place between the national elections of 1981 and the provincial elections in 1982. Whereas in 1981 party component 1(2) was related to ward component 2(1), in the 1982 elections the situation is completely reversed. To see what this means we will produce a combined joint plot for the three elections (Fig. 15.5). In this plot the most influential parties are shown as directions in the ward space. The directions were produced by postmultiplying the party space with each core planes in turn, and interpeting the resulting coordinates as vectors in the ward space.

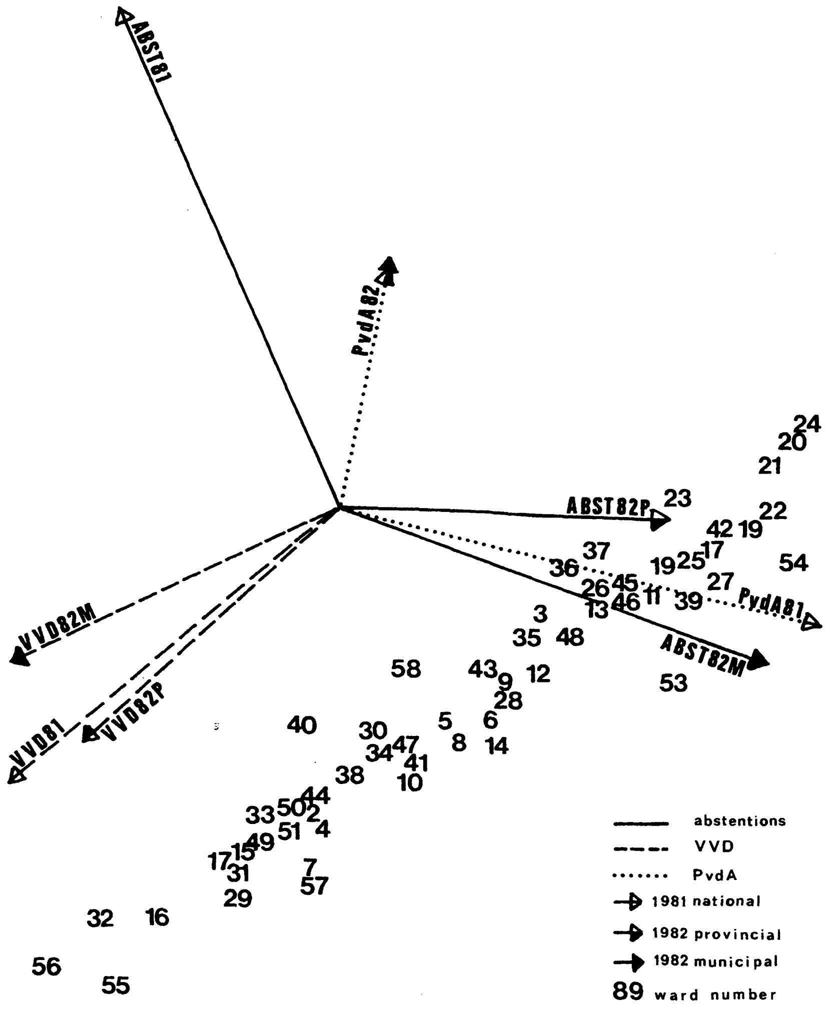


Fig. 15.5 *Leiden electorate study: Model I - changing party alliances*

The conclusions from Model I are the same as those from Models II and III combined, but the graphical representation in Fig. 15.5 makes it easier to formulate certain conclusions. It can be seen that the VVD has a stable electorate, e.g. in wards 56, 33, 55, 16, 1 etc. the VVD scored above expectation in all three elections, and in 20, 24, 21, etc. it scored below expectation.

Abstention in the national elections of 1981 is almost unrelated with party preference in the wards for the VVD and PvdA, as its direction in 1981 is nearly orthogonal to the main axis of the wards. In addition, abstention was below expectation. This reflects the very high abstention rate in the other two elections; after all the interaction effects should sum to zero.

Abstention in the provincial elections of 1982 is very strongly related to PvdA-voting in 1981. In other words, a high proportion of the wards which scored above expectation for the PvdA in 1981, abstained above expectation for the provincial elections in 1982. The same is true for the municipal elections, only less so. There is a considerable projection of the absentions in the municipal elections in the direction perpendicular to the main ward axis, indicating that also for other than PvdA wards abstention was above expectation in that election, but unrelated to the party preference of the wards. Note, by the way, that no statements can be made in terms of individual voters, as we have only the data available aggregated at the ward level.

The projections of all wards on the common PvdA82 direction are small, but generally higher for those wards which voted PvdA in 1981. Note, however, that only thirteen wards scored above expectation on this vector, so that the most stable support for the PvdA can be found there.

15.9 CONCLUSION

From the above discussion and example it can be seen that three-mode principal component analysis can be fruitfully used to study interactions in large three-way contingency tables. At the same time it is clear that it is helpful, and may be essential, to precede the analysis by a loglinear analysis to gain insight in the importance of the various interaction effects, and to decide for which model the chi-terms have to be evaluated. Without such insight it is difficult to attribute the observed interactions to particular effects.

In a sense the example is easier to interpret than data with a sizeable three-way interaction. On the other hand, particularly in the case of three-way interactions a three-mode correspondence analysis has much to offer that is not easily available in other techniques.

ACKNOWLEDGEMENTS

Chapter 1 :	The larger part of section 1.4 will also appear in Kroonenberg (1983, in press).
Chapter 2 :	Sections 2.1-2.4 will also appear in Kroonenberg (forthcoming); sections 2.5-2.10 are based on Kroonenberg (1981a).
Chapter 4 :	Sections 4.3-4.6 are based on Kroonenberg & De Leeuw (1980) reproduced by permission of the Psychometric Society.
Chapter 6 :	Sections 6.1-6.6 have been submitted for publication; Section 6.8 will also appear in Kroonenberg (forthcoming) in a slightly revised form.
Chapter 7 :	The data used in this chapter were collected within the framework of the project on TV -violence "Kritische waarneming van TV-geweld" (SVO-project 0543; principal investigator Dr. T.H.A. van der Voort). This project was supported financially by the 'Stichting voor Onderzoek van het Onderwijs (SVO)' and the 'Nederlandse Omroep Stichting (NOS)'.
Chapter 8 :	The material in this chapter will also appear as part of Kroonenberg (forthcoming).
Chapter 9 :	This chapter contains material from Osgood & Luria (1954) reproduced by permission of the American Psychological Association.
Chapter 10 :	This chapter contains material from Van der Kloot & Kroonenberg (1982), reproduced by permission of the editor of *Multivariate Behavioral Research*.
Chapter 13 :	The data for the *Hospital Study* were collected and prepared for analysis by Ms. Drs. W. Docter-de Leeuw within the framework of the project "Onderzoek naar de relatie tussen de groei van de ziekenhuisorgani-

	satie en de ontwikkeling van de directiestructuur en personeelssamenstelling" (principal investigators: prof. dr. C.J. Lammers and prof. dr. H. Philipsen). This project (nr. 50-6) was supported financially by the 'Nederlandse Stichting voor Zuiver Wetenschappelijk Onderzoek (ZWO)'.
Chapter 15 :	I am very grateful to Mr. J. Verhoeven, registered estate agent in Leiden, who was so kind to supply the estimates for Fig. 15.3.
Appendices :	The appendices will also appear in Kroonenberg(1983, in press).

APPENDICES

A.1 CLASSIFICATION OF THEORETICAL THREE-MODE PAPERS

Primarily three-mode analysis

ALS (PLS) - Kroonenberg & De Leeuw (1977,1978, 1980), Lohmöller & Wold (1980), Sands (1978), Sands & Young (1980).

Constrained T3 - Carroll, Pruzansky & Kruskal (1980).

Covariance structure approach - Bloxom (1968), Bentler & Lee (1978, 1979).

Exposition of T3 - Van de Geer (1975), Levin (1965), Lohmöller (1979a), Hohn (1979).

Rotation of core matrix - Cohen (1974,1975), MacCallum (1974a, 1976b), De Leeuw & Pruzansky (1978), Kroonenberg & De Leeuw (1977).

Scaling of input data - Kroonenberg (1981b).

T2-model - Israelsson (1969), Jennrich (1972), Carroll & Chang (1972), Kroonenberg & De Leeuw (1977,1978,1980), Tucker (1975).

T3-model - Tucker (1964,1965,1966a,1975), Levin (1963, 1965), Bartussek (1973), Jaffrennou (1978), Kroonenberg & De Leeuw (1980).

Three-mode scaling - Tucker (1972a,b, 1975).

Unique variances - F.W. Snyder (1968).

Weighted model (ALSCOMP3) - Sands (1978), Sands & Young (1980).

Closely related models/methods

CANDECOMP - Carroll & Chang (1970), Harshman (1970).

CANDELINC - Carroll, Pruzansky & Kruskal (1980).

INDSCAL - Carroll & Chang (1970), De Leeuw & Pruzansky (1978), Jaffrennou (1978).

IDIOSCAL - Carroll & Chang (1970, 1972), De Leeuw & Pruzansky (1978).

PARAFAC - Harshman (1970, 1972a,b, 1976).

PINDIS - Lingoes & Borg (1978).

Point-of-view analysis - Tucker & Messick (1963).

Review - Carroll & Arabie (1980), Carroll & Wish (1974), Lohmöller & Wold (1980), Law & Snyder (1979).

Taxonomy - Carroll & Arabie (1980).

Three-mode path analysis - Lohmöller & Wold (1980).

Three-mode point-of-view - Tzeng & Landis (1978).

Three-way unfolding - DeSarbo (1978), DeSarbo & Carroll (1979, 1981).

Vaguely related models

Double principal component analysis - Bourouche & Dussaix (1975).

Extension of 'binary method of Faverge' - Karnas (1975).

A.2 CLASSIFICATION OF APPLICATIONS: SUBJECT MATTER

Advertising

Buying behavior - Belk (1979)
Effectiveness for specific groups - Vavra (1972)
Product perception - Vavra (1973)
Viewer perception of advertising - Lastovicka (1981)

Developmental psychology

Changes in inkblot technique factors - Witzke (1975)
Changes in semantic differential - Lilly (1965)

Education

Achievement concepts - Knobloch (1972)
Aviation students - F.W. Snyder (1969)

Computer assisted instruction - Moonen (1978)
Educational careers - Stoop (1980)
Media usage - Lohmöller & Oerter (1979)
Multiple-cue learning - Montanelli (1972)
Novelty - Bernstein & Wicker (1969)
Progress in school subjects - Lohmöller (1978,1979a,1981a), Lohmöller & Wold (1980), Hanke, Lohmöller & Mandl (1980)
Serial learning - Love & Tucker (1970)
Stressful university situations - Kjerulff & Wiggins (1976)
Task learning - Tucker (1965, 1967), Fruchter (1969)
Task solving strategies - Rowe (1979)

Evoked potentials; EEGs

Evoked potentials - Donchin et al. (1972)
Various basic aspects of EEGs - Bartussek (1980)
Activity situations and EEGs - Bartessek & Gräser (1980), Bartussek et al. (1972)
Personality factors sensu Eysenck - Rösler (1972, 1975)

Geology

Organic extracts and elements - Hohn (1979)
Cations - Hohn & Friberg (1979)

Geography

Changes in land use - Bearwald (1976)
Changes in location of manufacturing - Cant (1971)
Spatial-temporal analysis - Chojnicki & Czyż (1976)

Law

Juvenile delinquents - Meijs (1980)

Occupational and organizational psychology; business administration

Administrative tasks - F.W. Snyder (1969)
Airline reservation agents - Inn, Hulin & Tucker (1972)
Hospital organization - Lammers (1974), Van de Geer (1974)
Job classification - Cornelius, Hakel & Sackett (1979)
Job satisfaction - Algera (1980), Zenisek (1980)
Organizational behaviour - Frederiksen (1972), Frederiksen et al. (1972)

Personality and social psychology

Abstract paintings - Baltink (1968,1969), Frey (1973), Litt (1966)

Achievement concepts	- Knobloch (1972)
Anxiousness	- Levin (1965), Tucker (1965)
Assertiveness	- Firth & Snyder (1979), Leah, Law & Snyder (1979)
Disjunctive conceptual behavior	- Snyder (1970, 1976)
Functional relations	- Groves (1978)
Gift giving	- Belk (1979)
Implicit theories of personality	- Wiggins & Blackburn (1976), Van der Kloot & Kroonenberg (1982)
Life events	- Redfield & Stone (1979), Saile (1979), Gräser, Esser & Saile (1981)
Manual expression	- Gitin (1970)
Perception of social environment	- Triandis (1976, 1977), Triandis et al. (1967, 1975)
Person stimuli	- Davis & Grobstein (1966)
Personality trait profiles	- Stewart (1971,1974)
Personality traits	- Schmitt, Coyle & Saari (1977)
Reversible figures	- Gräser (1977)
Self-conception	- Tzeng (1977b)
Self-report/peer-report	- Bentler & Lee (1978)
Social judgment	- Hirshberg (1980)
Social perception	- Imada & London (1979)
Social structure	- MacCallum (1974b)
Subjective culture	- Triandis (1972)

Phonetics

Confusion of consonants	- Kroonenberg & De Leeuw (1980)

Politics

American	- Sands (1978), Sands & Young (1980), Shikiar (1974a,b)
Dutch	- Kroonenberg (1981a,c), Kroonenberg & De Leeuw (1977,1978,1980)
German	- Rösler (1979)
Swedish	- Sjöberg (1977)
US Senate	- Wainer, Gruveaus & Zill (1973)

Psychiatry

Heart conditions	- Walter (1976)
Neuroticism	- Leichner (1975)
Schizophrenicity	- Mills & Tucker (1966)

Psychophysics

Psychomotor learning	- Fruchter (1969), Tucker (1965,1967)
Size-weight illusion	- Groves (1978)
Synesthetic thinking	- Wicker (1966,1968)
Sound quality	- Gabrielsson & Sjögren (1974/75)

Religion

Religious attitudes - Muthén et al. (1977)

Semantic differential studies

Affective meaning systems - Snyder & Wiggins (1970)
Affective and denotative meaning - Tzeng (1972,1975,1977a)
Cross-cultural - Tzeng & Landis (1978)
Thesaurus - Snyder (1967)
Self-concept description - Hentschel & Klintman (1974)
Developmental changes - Lilly (1965)

Stimulus scaling

Adjective similarity - Tucker (1972), MacCallum (1976b)
Confusions of consonants - Kroonenberg & De Leeuw (1980)
Personality traits - Van der Kloot & Kroonenberg (1982)
Soft drinks - Cooper (1973)
Sound quality - Gabrielsson & Sjögren (1974/75)

Various psychology

Adjective similarity - Tucker (1972a), MacCallum (1976b)
Leisure - London, Crandall & Fitzgibbons (1977)
Road research - Snyder & Law (1981)
Soft drinks - Cooper (1973)
Word association - Rychlak, Flynn & Burger (1979)

A.3 CLASSIFICATION OF APPLICATIONS: DATA TYPES

Semantic (or behavioral) differential scales

Baltink (1968,1969), Bernstein & Wicker (1969), Davis & Grobstein (1966), Frey (1973), Gitin (1970), Hentschel & Klintman (1974), Imada & London (1979). Leichner (1975), Levin (1964,1965), Litt (1966), MacCallum (1976b), Meijs (1980), Muthén et al. (1977), Redfield & Stone (1979), Snyder, F.W. (1967), Snyder F.W. & Wiggins (1970), Triandis (1972, 1976), Triandis et al. (1967,1975), Tzeng (1972,1975, 1977), Tzeng & Landis (1978), Wicker (1966,1968), Wiggins & Blackburn (1976).

Multitrait multimethod matrices analysed with Tucker (1966a) Method III or Bentler & Lee (1978,1979).

Bentler & Lee (1978,1979), Firth & Snyder, (1979), Hoffman & Tucker (1964), Leah, Law & Snyder, C.W. (1979), Schmitt, Coyle & Saari (1977), Snyder, C.W. (1970,1976), Snyder, F.W. (1967), Tucker (1965, 1966a,1967).

Time series data

Baerwald (1976), Bourouche & Dussaix (1975), Cant (1971), Gräser (1977), Hanke, Lohmöller & Mandl (1980), Inn, Hulin & Tucker (1972), Lammers (1974), Lohmöller (1978,1979a,1981a), Lohmöller & Wold (1980), Love & Tucker (1970), Van de Geer (1974).

Similarity type data

Cooper (1973), Kroonenberg (1981a,c), Kroonenberg & De Leeuw (1977, 1978,1980), MacCallum (1976b), Rösler (1979),Shikiar (1974a,b), Tucker (1972a, b).

A.4 REFERENCES TO COMPUTER PROGRAMS

ALS/PLS - Kroonenberg (1981a,c), Kroonenberg & De Leeuw (1980), Lohmöller & Wold (1980), Sands (1978), Sands & Young (1980)

Analysis of Covariance approach - Bentler & Lee (1978, 1979)

Orlik's Summax method - Kohler (1980)

Three-mode scaling - SOUPAC (1973), Redfield (1978)

Tucker's (1966a) Method I - Gräser (1977), Kouwer (1967), Lohmöller (1979b), McCloskey & Jackson (1979), Redfield (1978), SOUPAC (1973), Teufel (1969), Van de Geer (1975), Walsh (1964), Walsh & Walsh (1976).

Tucker's (1966a) Method II - Gruvaeus, Wainer & Snyder (1971), Lohmöller (1979b), Redfield (1978), Van de Geer (1975), Wainer et al. (1973)

Tucker's (1966a) Method III - Lohmöller (1979b), Redfield (1978), C.W. Snyder & Law (1979), C.W. Snyder, Law & Pamment (1979), SOUPAC (1973), Zenisek (1978).

REFERENCES

Ainsworth, M.D.S., Blehar, M.C., Waters, E., & Wall, S. *Patterns of attachment. A psychological study of the strange situation.* Hillsdale, N.J.: Lawrence Erlbaum, 1978.

Algera, J.A. *Kenmerken van werk.* (doctoral thesis, University of Leiden). Leiden, The Netherlands: Author, 1980.

Anderson, T.W. The use of factor analysis in statistical analysis of multiple time series. *Psychometrika*, 1963, *28*, 1-25.

Asch, S.E. Forming impressions of personality. *Journal of Abnormal and Social Psychology*, 1946, *41*, 253-290.

Baerwald, T.J. The emergence of a new "downtown". *The Geographical Review*, 1976, *68*, 308-318.

Baltink, G.J.H. Differentieel-psychologisch onderzoek naar de beoordeling van abstrakte schilderijen m.b.v. driemodale faktoranalyse. Unpublished master thesis, Institute for General Psychology, University of Groningen, Groningen, The Netherlands, 1968.

Baltink, G.J.H. Driemodale faktoranalyse in een differentieel-psychologisch onderzoek naar de beoordeling van abstracte schilderijen. *Nederlands Tijdschrift voor de Psychologie*, 1969, *24*, 529-540.

Bargmann, R.E. Exploratory techniques involving artificial variables. In P.R. Krishnaiah (Ed.), *Multivariate analysis II*. New York: Academic Press, 1969.

Barnett, V., & Lewis, T. *Outliers in statistical data*. Chichester, U.K.: Wiley, 1978.

Bartussek, D. Zur Interpretation der Kernmatrix in der dreimodalen Faktorenanalyse von L.R. Tucker. *Psychologische Beiträge*, 1973, *15*, 169-184.

Bartussek, D. Die dreimodale Faktorenanalyse als Methode zur Bestimmung von EEG-Frequenzbändern. In St. Kubicki, W.M. Herrmann & G. Laudahn (Eds.), *Faktorenanalyse und Variablenbildung aus dem Elektro-enzephalogramm*. Stuttgart, FRG.: Gustav Fischer Verlag, 1980 (Pp.15-26).

Bartussek, D. & Gräser, H. Ergebnisse dreimodaler Faktorenanalysen von EEG-Frequenzspektren. In St. Kubicki, W.M. Herrmann & G. Laudahn (Eds.), *Faktorenanalyse und Variablenbildung aus dem Elektro-enzephalogramm*. Stuttgart, FRG.: Gustav Fischer Verlag, 1980 (Pp.79-87) [1978].

Bartussek, D., Pawlik, K., Rhenius, D. Eine Dimensionsanalyse des digital frequenzanalysierten EEG und sein Zusammenhang mit Persönlichkeitsvariablen. Paper presented at the 13th meeting of Experimental Psychologists, Graz, Austria, 1972.

Bellman, R.R. *Introduction to matrix analysis*. New York, McGraw-Hill, 1960.

Belk, R.W. An exploratory assessment of situational effects in buyer behavior. *Journal of Marketing Research,* 1974, *11,* 156-163.

Belk, R.W. Gift-giving behavior. In J.N. Sheth (Ed.), *Research in marketing* (Vol. 2). Greenwich, CT: JAI Press, Inc., 1979 (Pp. 95-126).

Bentler, P.M. Assessment of developmental factor change at the individual and group level. In J.R. Nesselroode & H.W. Reese (Eds.), *Life-span developmental psychology. Methodological issues*. New York: Academic Press, 1973 (Pp.145-174).

Bentler, P.M. Multistructure statistical model applied to factor analysis. *Multivariate Behavioral Research*, 1976, *11*, 3-25.

Bentler, P.M. The interdependence of theory, methodology and empirical data: Causal modeling as an approach to construct validation. In D.B. Kandell (Ed.), *Longitudinal Research on Drug Use*. New York: Wiley, 1978.

Bentler, P.M. Structural equation models in longitudinal research. In S.A. Mednick & M. Harway (Eds.), *Longitudinal research in the United States*, 1980.

Bentler, P.M. & Lee, S.Y. Statistical aspects of three-mode factor analysis model. *Psychometrika,* 1978, *43,* 343-352.

Bentler, P.M. & Lee, S.Y. A statistical development of three-mode factor analysis. *British Journal of Mathematical & Statistical Psychology,* 1979, *32,* 87-104.

Bentler, P.M. & Weeks, D.G. Restricted multidimensional scaling models. *Journal of Mathematical Psychology*, 1978, *17*, 138-151.

Benzécri, J.P., & Collaborateurs. *L'analyse des données. II L'analyse des correspondances* (2nd edition). Paris, France: Dounod, 1976.

Bernstein, A.L. & Wicker, F.W. A three-mode factor analysis of the concept of novelty. *Psychonomic Science,* 1969, *14*, 291-292.

Bishop, Y.Y.M., Fienberg, S.E., & Holland, P.W. *Discrete multivariate analysis: Theory and practice*. Cambridge, Mass: The MIT Press, 1975.

Bloxom, B. A note on invariance in three-mode factor analysis. *Psychometrika,* 1968, *33,* 347-350 (a).

Bloxom, B. Individual differences in multidimensional scaling. ETS Research Bulletin 68-45, Educational Testing Service, Princeton, N.J., 1968 (b).

Bloxom, B. Constraint multidimensional scaling in N spaces. *Psychometrika*, 1978, *43*, 397-408.

Bloxom, B. Tucker's three-mode factor analysis model. In H.G. Law, C.W. Snyder, Jr., J. Hattie, & R.P. McDonald (Eds.), *Research methods for multi-mode data analysis*. New York: Praeger (forthcoming).

Bock, R.D. *Multivariate statistical methods in behavioral research*. New York: McGraw-Hill, 1975.

Borg, I. Facetten und Radextheorie in der multidimensionalen Skalierung. *Zeitschrift für Sozialpsychologie,* 1976, 7, 231-247.

Borg, I. & Lingoes, J.C. A model and algorithm for multidimensional scaling with external constraints on the distances. *Psychometrika*, 1980, *45*, 25-38.

Bouroche, J.M. & Dussaix, A.M. Several alternatives for three-way data analysis. *Metra,* 1975, *14,* 299-319.

Bruner, J.S. & Tagiuri, R. The perception of people. In G. Lindzey (Ed.), *Handbook of social psychology* (Vol.2). Cambridge, Mass: Addison-Wesley, 1954.

Bus, A.G. A longitudinal study in learning to read. Paper presented at the 9th World Congress on Reading, July 26-30, Dublin, Ireland, 1982.

Bus, A.G. & Kroonenberg, P.M. Reading instruction and learning to read: A longitudinal study. SOL-reeks, SOL/82-08, Department of Education, University of Groningen, Groningen, The Netherlands, 1982 (submitted for publication).

Cant, R.G., Changes in the location of manufacturing in New Zealand 1957-1968: An application of three-mode factor analysis. *New Zealand Geographer,* 1971, *27,* 38-55.

Carroll, J.B. Biquartumin criterion for rotation to oblique simple structure in factor analysis. *Science*, 1957, *126,* 1114-1115.

Carroll, J.D. Individual differences and multidimensional scaling. In R.N. Shepard, A.K. Romney, & S.B. Nerlove (Eds.), *Multidimensional scaling. Theory and applications in behavioral sciences*. New York: Seminar Press, 1972 (Pp.105-155).

Carroll, J.D., Application of CANDECOMP to solving for parameters of Lazarsfeld's latent class model. Presented at meeting of Society for Multivariate Experimental Psychology, Glenedan Beach, Oregon, November 13-15, 1975.

Carroll, J.D. & Arabie, P., Multidimensional scaling. *Annual Review of Psychology*, 1980, *31*, 607-649.

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

Carroll, J.D. & Chang, J.J. 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, New Jersey, March 30-31, 1972.

Carroll, J.D., Pruzansky, S., & Green, P.E. Estimation of the parameters of Lazarsfeld's latent class model of application of canonical decomposition (CANDECOMP) to multi-way contingency tables. Technical Memorandum 77-1229-7, Bell Laboratories, Murray Hill, NJ, 1977.

Carroll, J.D., Pruzansky, S., & Kruskal, J.B., CANDELINC: a general approach to multidimensional analysis of many-way arrays with linear constraints on parameters. *Psychometrika*, 1980, *45*, 3-24.

Carroll, J. D. & Wish, M. Models and methods for three-way multidimensional scaling . In D.H. Krantz, R.C. Atkinson, R.D. Luce, P. Suppes (Eds.), *Contemporary developments in mathematical psychology,* (Vol. II). San Francisco: W.H. Freeman & Cy., 1974 (Pp. 57-105).

Cattell, R.B. "Parallel proportional profiles" and other principles for determining the choice of factors by rotation. *Psychometrika*, 1944, *9*, 267-283.

Cattell, R.B. The data box: Its ordering of total responses in terms of possible relational systems. In R.B. Cattell (Ed.), *Handbook of multivariate experimental psychology*, Chicago, IL: Rand McNally, 1966(a) (Pp.67-128).

Cattell, R.B. Patterns of change: measurement in relation to state-dimension, trait change, lability, and process concept. In R.B. Cattell (Ed.), *Handbook of multivariate experimental psychology*, Chicago, IL: Rand McNally, 1966(b) (Pp.355-402).

Chatfield, C. *The analysis of time series: Theory and practice*. London, U.K.: Chapman and Hall, 1975.

Chojnicki, Z. & Czyż, T. Some problems in the application of factor analysis in geography. *Geographical Analysis,* 1976, *8,* 416-427.

Cliff, N. The 'idealized individual' interpretation of individual differences in multidimensional scaling. *Psychometrika*, 1968, *33*, 225-232.

Cohen, H.S. Three-mode rotation to approximate INDSCAL structure (TRIAS). Paper presented at Psychometric Society Meeting, Palo Alto,CA, 1974.

Cohen, H.S. Further thoughts on three-mode rotation to INDSCAL structure, with jackknifed confidence regions for points. Paper presented at U.S.-Japan Seminar on Theory, Methods and Applications of Multidimensional Scaling and Related Techniques. La Jolla, CA, 1975.

Cook, T.D., & Campbell, D.T. *Quasi-experimentation. Design and analysis issues for field settings*. Chicago, IL: Rand McNally, 1979.

Cooper, L.G. A multivariate investigation of preferences. *Multivariate Behavioral Research,* 1973, *8,* 253-272.

Corballis, M.C. Comparison of ranks of cross-product and covariance solutions in component analysis. *Psychometrika*, 1971, *36*, 243-249.

Corballis, M.C. A factor model for analysing change. *British Journal of Mathematical and Statistical Psychology*, 1973, *26*, 90-97.

Cornelius III, E.T., Hakel, M.D. & Sackett, P.R. A methodological approach to job classification for performance appraisal purposes. *Personnel Psychology,* 1979, *32,* 283-297.

Corsten, L.C.A. & Van Eijnsbergen, A.C. Multiplicative effects in two-way analysis of variance. *Statistica Neerlandica*, 1972, *26*, 61-68. (Corr. 27, 51)

Davis, E.E. & Grobstein, N.N. Multimode factor analysis of interpersonal perceptions. Technical Report No. 36, Department of Psychology, University of Illinois, Urbana-Champaign, IL, 1966.

De Gruijter, D. Cognitive structure of Dutch political parties. Report E091/67, Psychological Institute, University of Leiden, Leiden, The Netherlands, 1967.

De Leeuw, J. Generalized eigenvalue problems with positive semidefinite matrices. *Psychometrika*, 1982, *47*, 87-93.

De Leeuw, J. & Heiser, W. Multidimensional scaling with restrictions on the configuration. In P.R. Krishnaiah (Ed.), *Multivariate analysis-V*, Amsterdam, The Netherlands: North Holland, 1980 (Pp.501-522).

De Leeuw, J. & Pruzansky, S. A new computational method to fit the weighted Euclidean distance model. *Psychometrika,* 1978, *43,* 479-490.

DeSarbo, W.S. Three-way unfolding and situational dependence in consumer preference analysis. Unpublished doctoral thesis, University of Pennsylvania, Philadelphia, PA, 1978.

DeSarbo, W.S. & Carroll, J.D. Three-way unfolding. Unpublished working paper, Bell Laboratories, Murray Hill, NJ, 1979.

DeSarbo, W.S. & Carroll, J.D. Three-way unfolding and situational dependence in consumer preference analysis. Unpublished manuscript, Bell Laboratories, Murray Hill, NJ, 1981.

d'Esopo, D.A. A convex programming procedure. *Naval Research Logistic Quarterly*, 1959, *11*, 33-42.

Dixon, W.J. (Ed.) *BMDP statistical software 1981*. Berkeley, CA: University of California Press, 1981.

Donchin, E., Gerbrandt, L.A., Leifer, L. & Tucker, L.R. Is the contingent negative variation contingent on a motor response? *Psychophysiology*, 1972, *9*, 178-188.

Dunn, T.R. & Harshmann, R.A. A multidimensional scaling model for the size-weight illusion. *Psychometrika*, 1982, *47*, 25-45.

Dijkstra, T.K. *Latent variables in linear stochastic models. Reflections on "maximum likelihood" and "partial least squares" methods*. (Doctoral thesis, University of Groningen). Groningen, The Netherlands: Author, 1981.

Einhorn, H.J. Expert measurement and mechanical combination. *Organizational Behaviour and Human Performance*, 1972, 7, 86-106.

Firth, P.M. & Snyder Jr., C.W. Three-mode factor analysis of self-reported difficulty in assertiveness. *Australian Journal of Psychology*, 1979, *31*, 125-135.

Fischer, G.H. Zur Tuckers Methode der Faktorenanalyse von Lerndaten. *Psychologische Beiträge*, 1967, *10*, 135-146.

Fisher, R.A. & Mackenzie, W.A. Studies in crop variation II. The manurial response to different potato varieties. *Journal of Agricultural Science*, 1923, *13*, 311-320.

Frederiksen, N. Towards a taxonomy of situations. *American Psychologist*, 1972, *27*, 114-123.

Frederiksen, N., Jensen, O. & Beaton, A.E. *Prediction of organizational behavior*. Elmsford NY: Pergamon Press, 1972.

Frey, C. Profilskalierung von Kunstzeichnungen in Abhängigkeit von Persönlichkeitseigenschaften der Zeichner sowie der Beurteiler. Unpublished master thesis, University of Hamburg, FRG., 1973.

Fruchter, B. A comparison of two-mode and three-mode factor analysis of psychomotor learning performance. In H.R. Wijngaarden

(Pres.), *Proceedings of the XVIth International Congress of Applied Psychology* (Amsterdam, 18-22 August 1968). Amsterdam: Swets & Zeitlinger, 1969 (Pp.330-354).

Gabriel, K.R. The biplot graphic display of matrices with applications to principal components. *Biometrika*, 1971, *58*, 453-467.

Gabrielsson, A. Dimension analysis of perceived sound quality of sound-reproducing systems. *Scandinavian Journal of Psychology*, 1979, *20*, 159-169.

Gabrielsson, A. & Sjögren, H. Adjective ratings and dimension analysis of perceived sound quality of hearing aids I & II Report TA No. 75 & 77, Karolinska Institute, Technical Audiology, Stockholm, Sweden 1974 & 1975.

Gifi, A. *Nonlinear multivariate analysis*, Leiden, The Netherlands: Department of Data Theory, University of Leiden, 1981 (preliminary edition).

Gifi, A. Anacor user's guide. Department of Data Theory, University of Leiden, The Netherlands, 1982.

Gilbert, N. Non-additive combining of abilities. *Genetical Research Cambridge*, 1963, *4*, 213-219.

Gitin, S.R., A dimensional analysis of manual expression. *Journal of Personality and Social Psychology,* 1970, *15,* 271-277 (also unpublished master thesis, University of Illinois, 1968).

Glass, G.V. Primary, secondary, and meta-analysis of research. *Educational Researcher*, 1976, *5*, 3-8.

Glass, G.V., Wilson, V.L., & Gottman, J.M. *Design and analysis of time-series experiments*. Boulder, Colo.: Colorado Associated University Press, 1975.

Gnanadesikan, R. *Methods of statistical data analysis of multivariate observations*. New York: Wiley, 1977.

Gnanadesikan, R. & Kettenring, J.R. Robust estimates, residuals, and outliers with multiresponse data. *Biometrics*, 1972, *28*, 81-124.

Gollob, H.F. A statistical model which combines features of factor analysis and analysis of variance. *Psychometrika*, 1968, *33*, 73-115 (a).

Gollob, H.F. Confounding sources of variation in factor-analytic techniques. *Psychological Bulletin*, 1968, *70*, 330-344 (b).

Gollob, H.F. Rejoinder to Tucker's "Comments on confounding of sources of variation in factor-analytic techniques". *Psychological Bulletin*, 1968, *70*, 355-360 (c).

Goossens, F.A., Swaan, J., Tavecchio, L.W.C., Vergeer, M.M., & Van IJzendoorn, M.H. The quality of attachment assessed. WEP-reeks, WR 82-23-LE, Vakgroep W.E.P. University of Leiden, Leiden, The Netherlands, 1982.

Gower, J.C. Generalized procrustes analysis. *Psychometrika*, 1975, *40*, 33-51.

Gower, J.C. The analysis of three-way grids. In P. Slater (Ed.), *Dimensions of interpersonal space*. New York: Wiley, 1977.

Gräser, H., *Spontane Reversionsprozesse in der Figuralwahrnehmung. Eine Untersuchung reversibler Figuren mit der Dreimodalen Faktorenanalyse* (doctoral thesis, University of Trier). Trier, FRG: author, 1977.

Gräser, H., Esser, H. & Saile, H. Einschätzung von Lebensereignissen und ihren Auswirkungen. In S.H. Filipp (Ed.), *Kritische Lebensereignisse und ihre Bewältigung*. München, FRG: Urban & Schwarzenberg, 1981 (Pp. 104-122).

Green, P.E., Carmone, F.J., & Wachspress, D.P. Consumer segmentation via latent class analysis. *Journal of Consumer Research*, 1976, *3*, 170-174.

Gregory, R.T. & Karney, D.L. *A collection of matrices for testing computational algorithms*. New York: Wileu-Interscience, 1969.

Grossman, K.E., Grossman, K., Huber, F., & Wartner, K. German children's behavior towards their mother at 12 months and their father at 18 months in Ainsworth's Strange Situation. *International Journal of Behavioral Development*, 1981, *4*, 157-181.

Groves, C.L., Individual difference modelling of simple functional relations: Examples using three-mode factor analysis. Unpublished doctoral thesis, University of Illinois, Urbana-Champaign, IL, 1978. (*Dissertation Abstracts International*, 1978, 39 (5-B), 2475-2476).

Gruvaeus, G., Wainer, H., & Snyder, F. TREMOD: A 360/75 FORTRAN IV program for three mode factor analysis. *Behavioral Science*, 1971, *16*, 421-422.

Gruvaeus, G., Wainer, H., & Zill, N. Mixed modal matrices as aids to interpretation in 3-mode factor analysis. Paper presented at Psychometric Society Meeting, St. Louis, Mo., April, 1971.

Guttman, L. A new approach to factor analysis: The radex. In P.F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences*. New York: Free Press, 1954.

Haan, N., Common dimensions of personality development: Early adolescence in middle life. In D.H. Eichorn, P.H. Mussen, J.A. Clausen, N. Haan, & M.P. Honzik (Eds.), *Present and past in middle life*. New York: Academic Press, 1981 (Pp.117-151).

Haberman, S.J. Generalized residuals for log-linear models. *Proceedings of the 9th International Biometrics Conference*, 1976, *1*, 104-122.

Hakstian, A.R. Procedures for the factor analytic treatment of measures obtained on different occasions. *British Journal of Mathematical and Statistical Psychology*, 1973, *26*, 219-229.

Hamel, R.B. & Netelenbos, J.B. 'Structure d'ensemble': Eine Längsschnitt untersuchung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 1976, *8*, 263-273.

Hanke, B., Lohmöller, J.B., & Mandl, H. *Schülerbeurteilung in der Grundschule: Ergebnisse der Augsburger Längsschnittuntersuchung*. München, FRG.: Oldenbourg Verlag, 1980.

Harris, C.W. Relations among factors of raw deviation, and double-centered score matrices. *Journal of Experimental Education*, 1953, *22*, 53-58.

Harshman, R.A. Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-mode factor analysis. *UCLA Working Papers in Phonetics*, 1970, *16*, 1-84. (Reprinted by Xerox University Microfilms, Ann Arbor, Mi.; order no. 10,085)

Harshman, R.A. PARAFAC2: Mathematical and technical notes. *UCLA Working Papers in Phonetics*, 1972, *22*, 31-44(a). (Reprinted by Xerox University Microfilms, Ann Arbor, Mi.; order no.10,085)

Harshman, R.A. Determination and proof of minimum uniqueness conditions for PARAFAC1. *UCLA Working Papers in Phonetics*, 1972, *22*, 111-117(b). (Reprinted by Xerox University Microfilms, Ann Arbor, Mi.; order no. 10,085)

Harshman, R.A. PARAFAC: Methods of three-way factor analysis and multidimensional scaling according to the principle of proportional profiles. Unpublished doctoral thesis, University of California, Los Angeles, 1976. (*Dissertation Abstracts International*, 1976, *37*, (5-B), 2478-2479).

Harshman, R.A. & Berenbaum, S.A. Basic concepts underlying the PARAFAC-CANDECOMP three-way factor analysis and its application to longitudinal data. In D.H. Eichorn, P.H. Mussen, J.A. Clausen, N. Haan, & M.P. Honzik (Eds.), *Present and past in middle life*, New York: Academic Press, 1981.

Hawkins, D.M. The detection of errors in multivariate data using principal components. *Journal of the American Statistical Association*, 1974, *69*, 340-344.

Hawkins, D.M. *Identification of outliers*. London: Chapman and Hall, 1980.

Heise, D.R. Some methodological issues in semantic differential research. *Psychological Bulletin*, 1969, *72*, 406-422.

Hentschel, U. & Klintman, H. A 28-variable semantic differential. I. On the factorial identification of content. *Psychological Research Bulletin* (Lund University, Sweden), 1974, *14*(4), 1-27.

Hill, M.O. Correspondence analysis: a neglected multivariate method. *Journal of the Royal Statistical Society, Series C*, 1974, *23*, 340-354.

Hirschberg, N. Individual differences in social judgment: a multivariate approach. In M. Fishbein (Ed.), *Progress in social psychology*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1980.

Hoffman, E.L. & Tucker, L.R. Three-way factor analysis of a multitrait-multimethod matrix. Technical Report, Department of Psychology, University of Illinois, Urbana, IL, 1964.

Hohn, M.E. Principal component analysis of three-way tables. *Journal of the International Association of Mathematical Geology,* 1979, *11,* 611-626.

Hohn, M.E. & Friberg, L.M. A generalized principal components model in petrology. *Lithos,* 1979, *12,* 317-324.

Horan, C.B. Multidimensional scaling: Combining observations when individuals have different perceptual structures. *Psychometrika*, 1969, *34*, 139-165.

Householder, A.S. *The theory of matrices in numerical analysis.* New York: Blaisdell, 1964.

Hull, C.H. & Nie, N.H. *SPSS Update 7-9*. New York: McGraw-Hill, 1981.

Hurley, J.R. & Cattell, R.B. Procrustes program: Producing direct rotation to test the hypothesized factor structure. *Behavioral Science*, 1962, 7, 258-261.

Imada, A.S. & London, M. Relationships between subjects, scales and stimuli in research on social perception. *Perceptual and Motor Skills,* 1979, *48,* 691-697.

Inn, A., Hulin, C.L., & Tucker, L.R. Three sources of criterion variance: static dimensionality, dynamic dimensionality, and individual dimensionality. *Organizational Behavior and Human Performance,* 1972, *8,* 58-83.

Israels, A.Z., Bethlehem, J.G., Van Driel, J., Jansen, M.E., Pannekoek, J., De Ree, S.J.M., & Sikkel, D. Multivariate analysemethoden voor discrete variabelen. *Kwantitatieve methoden*, 1981, 2(2), 87-149.

Israelsson, A., Three-way (or second order) component analysis. In H. Wold & E. Lyttkens (Eds.), Nonlinear iterative partial least- squares (NIPALS) estimation procedures. *Bulletin of the International Statistical Institute,* 1969, *43,* 29-51.

Jackson, J.E. Principal component analysis and factor analysis: Part I-Principal components. *Journal of Quality Technology*, 1980, *12*, 201-213.

Jackson, J.E. & Bradley, R.A. Sequential multivariate procedures for means with quality control applications. In P.R. Krisnaiah (Ed.), *Multivariate analysis*, New York: Academic Press, 1966 (Pp.507-519).

Jackson, J.E. & Morris, R.H. An application of multivariate quality control to photographic processing. *Journal of the American Statistical Association*, 1957, *52*, 186-199.

Jackson, J.E. & Mudholkar, G.S. Control procedures for residuals associated with principal component analysis. *Technometrics*, 1979, *21*, 341-349.

Jaffrennou, P.A. Sur l'analyse des familles finies de variables vectorielles. Bases algèbriques et application à la description statistique. Pré-publication No. 4. Département de Mathématiques. University of Saint-Etienne, France, 1978.

Jansen, G.G.H. & Bus, A.G. An application of a logistic bio-assay model to achievement test data. Technical Report, Department of Education, University of Groningen, The Netherlands, 1982 (submitted for publication).

Jennrich, R. A generalization of the multidimensional scaling model of Carroll & Chang. *UCLA Working Papers in Phonetics,* 1972, *22*, 45-47.

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

Jones, L.E. & Young, F.W. Structure of social environment: Longitudinal individual differences scaling of an intact group. *Journal of Personality and Social Psychology*, 1972, *24*, 108-121.

Jöreskog, K.G. Simultaneous factor analysis in several populations. *Psychometrika*, 1971, *36*, 409-426.

Jöreskog, K.G. Analysis of covariance structures. In P.R. Krisnaiah (Ed.) *Multivariate analysis-III*. New York: Academic Press, 1973(a) (Pp.263-285).

Jöreskog, K.G. A general method for estimating a linear structural equation system. In A.S. Goldberger & O.D. Duncan (Eds.), *Structural equation models in the social sciences*. New York: Seminar Press, 1973(b) (Pp.85-112).

Jöreskog, K.G. Structural analysis of covariance and correlation matrices. *Psychometrika*, 1978, *43*, 443-477 (a).

Jöreskog, K.G. An economic model for multivariate panel data. *Annales de l'INSEE*, no. 30-31, 1978 (b).

Jöreskog, K.G. Statistical estimation of structural models in longitudinal developmental investigation. In J.R. Nesselroade & P.B. Baltes (Eds.), *Longitudinal methodology in the study of behavior and human development*. New York: Academic Press, 1979.

Jöreskog, K.G. & Sörbom, D. Statistical models and methods for analysis of longitudinal data. In D.J. Aigner & A.S. Goldberger (Eds.), *Latent variables in socio-economic models*. Amsterdam, The Netherlands: North Holland, 1977 (Pp.285-325).

Jöreskog, K.G. & Sörbom, D. LISREL IV. A general computer program for estimation of linear structural equation systems by maximum likelihood methods. User's guide. Department of Statistics, University of Uppsala, Sweden, 1978.

Kaiser, H.F. The varimax rotation for analytic rotation in factor analysis. *Psychometrika*, 1958, *23*, 187-200.

Kapteijn, A., Neudecker, H., & Wansbeek, T. N-mode component analysis. Technical report, Netherlands Central Bureau of Statistics, Voorburg, The Netherlands, 1982 (submitted for publication).

Karnas, G. Note sur une procédure d'analyse de données relatives à une correspondance ternaire ou pseudo-ternaire par la méthode d'analyse binaire de Faverge. *Le Travail Humain*, 1975, *38*, 287-300.

Kendall, M.G. *A course in multivariate analysis*. London, U.K.: Griffin, 1957 (1975).

Kjerulff, K. & Wiggins, N.H. Graduate student styles for coping with stressful situations. *Journal of Educational Psychology*, 1976, *68*, 247-254.

Knobloch, E.M. Einschätzung von leistungsrelevanten Begriffen. Unpublished master thesis, University of Hamburg, Hamburg, FRG., 1972.

Kohler, A., Das Trimod-Programm-System (TRIPSY) zur Berechnung der dreimodalen Faktorenanalyse nach Orlik, (manuscript in preparation).

Kouwer, B.J. Drie-modale faktoranalyse. Programmabeschrijving (GRON. PSYCH. 07+07BIS). Orthogonale rotaties (GRON.PSYCH.12). Reports, Institute of Psychology, University of Groningen, Groningen, The Netherlands, 1967.

Kouwer, B.J. & Hartong, N. Het scoren van een arbeidscurve. *Nederlands Tijdschrift voor de Psychologie*, 1961, *16*, 184-193.

Kroonenberg, P.M., User's Guide to TUCKALS3. A program for three-mode principal component analysis. WEP-reeks, WR 81-6-RP, Vakgroep W.E.P., University of Leiden, Leiden, The Netherlands, 1981, [1979](a).

Kroonenberg, P.M. Scaling of input data for three-mode principal component analysis. WEP-reeks, WR 81-21-EX, Vakgroep W.E.P., University of Leiden, Leiden, The Netherlands, 1981(b).

Kroonenberg, P.M., User's guide to TUCKALS2. A program for three-mode principal component analysis with extended core matrix. WEP-reeks, WR-81-35-RP, Vakgroep W.E.P., University of Leiden, Leiden, The Netherlands, 1981(c).

Kroonenberg, P.M. Annotated Bibliography of three-mode factor analysis. *British Journal of Mathematical and Statistical Psychology*, 1983, *36* (in press).

Kroonenberg, P.M. Three-mode principal component analysis illustrated with an example from attachment theory. In H.G. Law, C.W. Snyder Jr., J. Hattie, & R.P.McDonald (Eds.), *Research methods for multi-mode data analysis*. New York: Praeger (forthcoming).

Kroonenberg, P.M. & De Leeuw, J. TUCKALS2: A principal component analysis of three mode data. Res. Bull. RB. 001-77, Department of Data Theory, University of Leiden, Leiden, The Netherlands, 1977.

Kroonenberg, P.M. & De Leeuw, J. TUCKALS2: Een hoofdassenanalyse voor drieweggegevens. *Methoden en Data Nieuwsbrief* (vd SWS vd VVS), 1978, *3* (3), 30-53.

Kroonenberg, P.M. & De Leeuw, J. Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika,* 1980, *45,* 69-97.

Kroonenberg, P.M. & Lewis, C. Methodological issues in search of factor model: exploration through confirmation. *Journal of Educational Statistics*, 1982, 7, 69-89.

Kruskal, J.B. More factors than subjects, tests, and treatments: An indeterminacy theorem for canonical decomposition and individual difference scalings. *Psychometrika*, 1976, *41*, 281-293.

Kruskal, J.B. Three-way arrays: Rank and uniqueness of trilinear decompositions, with applications to arithmetic complexity and statistics. *Linear Algebra and its Applications*, 1977, *18*, 95-138.

Kruskal, J.B. Factor analysis and principal component analysis I. Bilinear methods. In W.H. Kruskal & J. Tenur (Eds.), *International encyclopedia of statistics*. San Francisco, CA: Freeman & Cy., 1987.

Kruskal, J.B. Multilinear models for data analysis. *Behavior metrika*, 1981, *10*, 1-20.

Kruskal, J.B. & Wish, M. *Multidimendional scaling*. (Sage University Paper, nr. 11). Beverly Hills, CA: Sage Publications, 1978.

Lammers, C.J. Groei en ontwikkeling van de ziekenhuisorganisaties in Nederland. Interimrapport, Institute of Sociology, University of Leiden, Leiden, The Netherlands, 1974.

Lastovicka, J.L. The extension of component analysis to four-mode matrices. *Psychometrika,* 1981, *46,* 47-57.

Law, H.G. & Snyder Jr. C.W. Three-mode models for the analysis of psychological data. *Australian Psychologist,* 1979, *14,* 214 (conference abstract).

Law, H.G. & Snyder Jr., C.W. An introduction to the analysis of covariance structures: a general model for data analysis. In J.M. Morris (Ed.), *Proceedings of a seminar on measuring social behaviour in road research.* Vermont South, Vic., Australia: Australian Road Research Board, 1981 (Pp. 49-60).

Law, H.G., Snyder, Jr., C.W., Hattie, J.P., & McDonald, R.P. (Eds.) *Research methods for multi-mode data analysis.* New York: Praeger, forthcoming.

Lazarsfeld, P.F., & Henry, .W. *Latent structure analysis.* Boston: Houghton Mifflin, 1968.

Leah, J.A., Law, H.G. & Snyder Jr., C.W. The structure of self-reported difficulty in assertiveness: An application of three-mode common factor analysis. *Multivariate Behavioral Research,* 1979, *14,* 443-462.

Lee, S.Y. & Fong, W.K. A modified model for three-mode factor analysis. Technical Report, The Chinese University of Hong Kong, 1982 (submitted for publication).

Leichner, R. Zur Verarbeitung psychiatrischer Information I. *Diagnostica,* 1975, *21,* 147-166.

Levin, J. Three-mode factor analysis. Unpublished doctoral thesis, University of Illinois, Urbana, IL, 1963. *(Dissertation Abstracts International,* 1964, *24* (12), 5530-5531).

Levin, J. Three-mode factor analysis. *Psychological Bulletin,* 1965, *64,* 442-452.

Lilly, R.S. A developmental study of the semantic differential. ETS Research Bulletin 65-28; Unpublished doctoral dissertation, Princeton University, Princeton, N.J., 1965. *(Dissertation Abstracts International,* 1966, *26,*(7), 4063-4064).

Lingoes, J.C. & Borg, I. A direct approach to individual differences scaling using increasingly complex transformations. *Psychometrika,* 1978, *43,* 491-520.

Linschoten, J. *Idolen van de psycholoog.* Utrecht, The Netherlands: Erven J. Bijleveld, 1964.

Litt, E.N. A factorial study of responses to abstract paintings. Unpublished master thesis, University of Illinois, Urbana, IL, 1966.

Lohmöller, J.B. How longitudinal factor stability, continuity, differentiation, and integration are portrayed 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.

Lohmöller, J.B., Die trimodale Faktorenanalyse von Tucker: Skalierungen, Rotationen, andere Modelle. *Archiv für Psychologie,* 1979, *131,* 137-166(a).

Lohmöller, J.B. Programmbeschreibung von FA-3-Trimodale Faktorenanalyse. In J.B. Lohmöller, Das COR-Programm-system zur Korrelationsanalyse. Fachbereich Pädagogik, Hochschule der Bundeswehr München, Neubiberg, FRG., 1979(b).

Lohmöller, J.B. Stabilität und Kontinuität in Längsschnittdaten, analysiert durch T- und trimodale Faktorenanalyse. Forschungsbericht, Fachbereich Pädagogik, Hochschule der Bundeswehr München, Neubiberg, FRG., 1981 [1978] (a).

Lohmöller, J.B. Pfadmodelle mit latenten Variabelen: LVPLSC ist eine leistungsfähige Alternative zu Lisrel. Forschungsbericht 81-02, Fachbereich Pädagogik, Hochschule der Bundeswehr München, Neubiberg, FRG., 1981 (b).

Lohmöller, J.B. & Oerter, R. (Eds.), *Medien in der Erzieherausbildung: Erprobung des Medienverbundes "Vorschulische Erziehung im Ausland".* München, FRG.: Oldenbourg Verlag, 1979 (Pp.114-130).

Lohmöller, J.B. & Wold, H. Three-mode path models with latent variables and partial least squares (PLS) parameter estimation. Forschungsbericht 80.03, Fachbereich Pädagogik, Hochschule der Bundeswehr München, Neubiberg, FRG., 1980; revised 1982 (presented at European Meeting of the Psychometric Society, Groningen, The Netherlands, June 18-21, 1980).

Lohmöller, J.B. & Wold, H. Pfad- und faktorenanalytische Ansätze zur differentiellen Entwicklungsbeschreibungen : Die trimodale Pfadanalyse mit latenten Variabelen. In R. Oerter (Ed.), *Bericht über die 5.Tagung Entwicklungspsychologie in Augsburg (Vol.I)*, 21-23 september, 1981 (Pp.36-43).

London, M., Crandall, R. & Fitzgibbons, D. The psychological structure of leisure: Activities, needs, people. *Journal of Leisure Research,* 1977, *9,* 252-263.

Love, W.D. & Tucker, L.R. A three-mode factor analysis of serial learning. Report of the Office of Naval Research, 1970.

MacCallum, R.C. A comparison of two individual differences models for multidimensional scaling: Carroll and Chang's INDSCAL and Tucker's three-mode factor analysis. Unpublished doctoral thesis, University of Illinois, Urbana, IL, 1974(a). *(Dissertation Abstracts International,* 1975, *35,* (7-B), 3619).

MacCallum, R.C. Relations between factor analysis and multidimensional scaling. *Psychogical Bulletin,* 1974, *81,* 505-516(b).

MacCallum, R.C. Effects on INDSCAL of non-orthogonal perceptions of object space dimensions. *Psychometrika,* 1976, *41,* 177-188 (a).

MacCallum, R.C. Transformation of a three-mode multidimensional scaling solution to INDSCAL form. *Psychometrika,* 1976, *41,* 385-400 (b).

Mandel, J. The partitioning of interaction in analysis of variance. *Journal of Research. National Bureau of Standards. Section B. Mathematical Sciences*, 1969, *73B*, 309-328.

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

Marasinghe, M.G. & Johnson, D.E. Testing subhypotheses in the multiplicative interaction model. *Technometrics*, 1981, *23*, 385-393.

McCall, R.B., Appelbaum, M. & Hogarty, P.S. Patterns of IQ change over age. *Monographs of the Society for Research in Child Development*, 1973, *38* (3).

McCloskey, J. & Jackson, P.R. THREE-MODE: A FORTRAN IV program for three-mode factor analysis. *Behavior Research Methods and Instrumentation,* 1979, *11,* 75-76.

McDermott, B.J. Multidimensional analysis of circuit quality judgments. *Journal of the Acoustical Society of America*, 1969, *45*, 774-781.

McDonald, R.P. Nonlinear factor analysis. *Psychometric Monographs*, 1967, *15*.

McDonald, R.P. A simple comprehensive model for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 1978, *31*, 59-72.

Meerling. *Methoden en technieken van psychologisch onderzoek.* Meppel, The Netherlands: Boom, 1980.

Meijs, B.W.G.Ph. Huis van bewaring en subkultuur: een empirische studie bij jeugdige gedetineerden naar het effekt van '102 dagen' preventieve hechtenis op attitudes en andere indikatoren van subkultuur. Unpublished master thesis, University of Leiden, The Netherlands, 1980.

Meredith, W. Rotation to achieve factorial invariance. *Psychometrika*, 1964, *29*, 187-206.

Meuwese, W. Een vergelijking van twee methoden van beoordeling van verbale stimuli. *Nederlands Tijdschrift voor de Psychologie*, 1970, *25*, 594-603.

Meyer, R.R. The validity of a family of optimization methods. *SIAM Journal of Control and Optimization*, 1970, *15*, 699-715.

Meyers, C.E., Dingman, H.F., Orpet, R.E., Sitkei, E.G., Watts, C. A. Four ability-factor hypotheses at three preliterate levels in normal and retarded children. *Monograph Society for Research in Child Development*, 1964, *29*(5), 1-80.

Mills, D.H. & Tucker, L.R. A three-mode factor analysis of clinical judgment of schizophrenicity. *Journal of Clinical Psychology*, 1966, *22*, 136-139.

Montanelli, D.S. Multiple-cue learning in children. *Developmental Psychology*. 1972, *7*, 302-312.

Moonen, J. *Computergestuurd Onderwijs*. Een onderzoek naar de mogelijkheden tot geïntegreerd gebruik van een computergestuurd systeem in een statistiekkurrikulum (doctoral thesis, University of Leiden). Leiden, The Netherlands: Author, 1978.

Morrison, D.F. *Multivariate statistical methods* (2nd edition). New York: McGraw-Hill, 1976.

Muthén, B., Olsson, U., Pettersson, T. & Stahlberg, G. Measuring religious attitudes using the semantic differential technique: An application of three-mode factor analysis. *Journal for the Scientific Study of Religion*, 1977, *16*, 275-288.

Nesselroade, J.A. Faktorenanalyse von Kreuzprodukten zur Beschreibung von Veränderungsphänomenen (Change). *Zeitschrift für experimentelle und angewandte Psychologie*, 1973, *20*, 92-106.

Nishisato, S. *Analysis of categorical data: Dual scaling and its applications*. Toronto, Canada: University of Toronto Press, 1980.

Noy-Meir, I. Data transformations in ecological ordination. I. Some advantages of non-centering. *Journal of Ecology*, 1973, *61*, 329-341.

Noy-Meir, 1., Walker, D., & Williams, W.T. Data transformations in ecological ordination: On the meaning of data standardization. *Journal of Ecology*, 1975, *63*, 779-800.

Orlik, P. Das Summax-Modell der dreimodalen Faktorenanalyse mit interpretierbarer Kernmatrix. Technical Report, University of Saarland, Saarbrücken, FRG., 1980 [1976].

Osgood, C.E. & Luria, Z. A blind analysis of a case of multiple personality. *Journal of Abnormal and Social Psychology*, 1954, *49*, 579-791. [Reprinted in J.G. Snider & C.E. Osgood (Eds.), *Semantic differential technique. A source book.* Chicago, IL: Aldine, 1969 (Pp.505-517)].

Osgood, C.E. & Suci, G.J. A measure of relation determined by both mean difference and profile information. *Psychological Bulletin*, 1952, *49*, 251-262.

Osgood, C.E., Suci, G.J. & Tannenbaum, P. *The measurement of meaning.* Urbana, IL: University of Illinois Press, 1957.

Ostrowski, A.M. *Solution of equations and systems of equations.* New York: Academic Press, 1966.

Penrose, R. On the best approximate solutions of linear matrix equations. *Proceedings of the Cambridge Philosophical Society*, 1955, *51*, 406-413.

Pruzansky, S. How to use SINDSCAL. A computer program for individual differences in multidimensional scaling. Technical Report, Bell Laboratories, Murray Hill, NJ, 1975.

Rao, C.R. The use and interpretation of principal component analysis in applied research. *Sankyā A*, 1964, *26*, 329-358.

Redfield, J. TMFA: A FORTRAN program for three-mode factor-analysis and individual-differences in multidimensional-scaling. *Educational and Psychological Measurement*, 1978, *38*, 793-795.

Redfield, J. & Stone, A. Individual view points of stressful life events. *Journal of Consulting and Clinical Psychology*, 1979, *47*, 147-154.

Roskam, E.E. Multivariate analysis of change and growth: Critical review and perspectives. In D.N.M. de Gruijter & L.J.T. van der Kamp (Eds.), *Advances in psychological and educational measurement.* London, U.K.: Wiley, 1976 (Pp.111-133).

Rösler, F. Dimensionen der Aktivitäten und deren Beziehungen zu den Persönlichkeitsfaktoren "Extraversion/Introversion" und "Neurotizismus" sensu Eysenk. Unpublished master thesis, University of Hamburg, Hamburg, FRG., 1972.

Rösler, F. Die Abhängigkeit des Elektroenzephalogramms von den Persönlichkeitsdimensionen E und N sensu Eysenk und unterschiedlich aktivierenden Situationen. *Zeitschrift für experimentelle und angewandte Psychologie*, 1975, *22*, 630-667.

Rösler, F. Identifying interindividual judgment differences: INDSCAL or three-mode factor analysis. *Multivariate Behavioral Research*, 1979, *14*, 145-167.

Ross, J. The relation between test and person factors. *Psychological Review*, 1963, *70*, 432-443.

Ross, J. Mean performance and factor analysis of learning data. *Psychometrika*, 1964, *29*, 67-73.

Ross, J. A remark on Tucker and Messick's "point of view" analysis. *Psychometrika*, 1966, *31*, 27-31.

Rowe, H.A.H. Three-mode factor analysis: problems of interpretation and possible solutions. *Australian Psychologist*, 1979, *14*, 222-223 (abstract of paper presented at the 14th Annual Conference of the Australian Psychological Society, University of Tasmania, 30 August, 1979).

Ruch, W. Gemeinsame Strukturen in Weizbeurteilung und Persönlichkeit. Versuch einer empirischen Integration des Gegenstandsbereiches Witzbeurteilung in die differentielle Psychologie. Unpublished doctoral thesis, University of Graz, Austria, 1980.

Ruch, W. Witzbeurteilung und Persönlichkeit: Ein trimodale analyse. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 1981, *2*, 253-273.

Rutishauser, H. Computational aspects of F.L. Bauer's simultaneous iteration method. *Numerische Mathematik*, 1969, *13*, 4-13.

Rychlak, J.F., Flynn, E.J. & Burger, G., Affection and evaluation as logical processes of meaningfulness independent of associative frequency. *Journal of General Psychology*, 1979, *100*, 143-157.

Saile, H. Zur Struktur der Einschätzung von Lebensereignissen. Eine Untersuching über Beurteilungsunterschiede mittels dreimodaler Faktorenanalyse. Unpublished master thesis: University of Trier, Trier, FRG., 1979.

Sands, R. Component models for three-way data: ALSCOMP3, an alternating least squares algorithm with optimal scaling features. Unpublished master thesis, Department of Psychology, University of North Carolina, Chapel Hill, N.C., 1978.

Sands, R. & Young, F.W. Component models for three-way data: ALSCOMP3, an alternative least squares algorithm with optimal scaling features. *Psychometrika*, 1980, *45*, 39-67 [1978].

Sawyer, J. Measurement and prediction, clinical and statistical. *Psychological Bulletin*, 1966, *66*, 178-200.

Schiffman, H. & Falkenberg, P. The organization of stimuli and sensory neurons. *Physiology and Behavior*, 1968, *3*, 197-201.

Schiffman, S.S., Reynolds, M.L., & Young, F.W. *Introduction to multidimensional scaling. Theory, methods, and applications.* New York: Academic Press, 1981.

Schmitt, N., Coyle, B.W. & Saari, B.B. A review and critique of analyses of multitrait-multimethod matrices. *Multivariate Behavioral Research,* 1977, *12,* 447-478.

Schneider, D.J. Implicit personality theory: A review. *Psychological Bulletin*, 1973, *79*, 294-309.

Schönemann, P.H. An algebraic solution for a class of subjective metrics models. *Psychometrika*, 1972, *37*, 3-27.

Schwartz, H.R., Rutishauser, H., & Stiefel, E. *Numerik. Symmetrischer Matrizen*. Stuttgart: Teubner, 1968.

Shepard, R.N. The circumplex and related topological manifolds in the study of perception. In S. Shye (Ed.), *Theory construction and data analysis in the behavioral sciences.* San Francisco: Jossey-Bass, 1978.

Shikiar, R. The perception of politicians and political issues: a multidimensional scaling approach. *Multivariate Behavioral Research,* 1974, *9,* 461-477(a).

Shikiar, R. An empirical comparison of two individual differences multidimensional scaling models. *Educational and Psychogical Measurement,* 1974, *34,* 823-828(b).

Sjöberg, L. Choice frequency and similarity. *Scandinavion Journal of Psychology,* 977, *18,* 103-115.

Skolnick, A. Married lives: longitudinal perspectives on marriage. In D.H. Eichorn, P.H. Mussen, J.A. Clausen, N. Haan, & M.P. Honzik (Eds.), *Present and past in middle life*. New York: Academic Press, 1981 (Pp.269-298).

Snee, R.D. On the analysis of response curve. *Technometrics*, 1972, *14*, 47-62.

Snee, R.D., Acuff, S.K., & Gibson, J.R. A useful method for the analysis of growth studies. *Biometrics,* 1979, *35*, 835-848.

Snider, J.G. & Osgood, C.E. (Eds.) *Semantic differential technique. A source book*. Chicago, IL: Aldine, 1969.

Snyder Jr., C.W. Intrinsic individual differences in disjunctive conceptual behavior: three-mode factor analysis. Unpublished doctoral thesis, University of Pennsylvania, 1970. (*Dissertation Abstracts International,* 1970, *32,* (1-B), 544.)

Snyder Jr., C.W. Multivariate analysis of intrinsic individual differences in disjunctive conceptual behavior. *Multivariate Behavioral Research,* 1976, *11,* 195-216.

Snyder Jr., C.W., Bridgman, R.P., & Law, H.G. Three-mode factor analytic reference curves for concept identification. *Personality and Individual Differences,* 1981, *2,* 265-272.

Snyder Jr., C.W. and Law, H.G. Three-mode common factor analysis: procedure and computer programs. *Multivariate Behavioral Research,* 1979, *14,* 435-441.

Snyder Jr., C.W., Law, H.G. & Pamment, P.R. Calculation of Tucker's three-mode common factor analysis. *Behavior Research Methods and Instrumentation,* 1979, *11,* 609-611.

Snyder, Jr. C.W. & Law, H.G. Three-mode models for road research. In: J.M. Morris (Ed.), *Proceedings of a Seminar on measuring social behaviour in road research.* Vermont South, Vic., Australia: Australian Road Research Board, 1981, (Pp. 39-48).

Snyder F.W. An investigation of the invariance of the semantic differential across the subject mode. Unpublished master thesis, University of Illinois, Urbana, IL,1967.

Snyder, F.W. A unique variance model for three-mode factor analysis. Department of Psychology, University of Illinois, Urbana, IL, 1968 (also doctoral thesis, 1969). (*Dissertation Abstracts International,* 1969, *30,* (3-B), 1349).

Snyder, F.W. & Tucker, L.R. On the interpretation of the core matrix in three-mode factor analysis. Paper read at the Psychometric Society Meeting, March, 1970.

Snyder, F.W. & Wiggins, N. Affective meaning systems: A multivariate approach. *Multivariate Behavioral Research,* 1970, *5,* 453-468 [1968].

SOUPAC program descriptions. Computing Services Offices, University of Illinois, Urbana, IL, 1973.

Sroufe, L.A. & Waters, E. Attachment as an organizational construct. *Child Development,* 1977, *48,* 1184-1199.

Stewart, T.R. The relation between three-mode factor analysis and multidimensional scaling of personality trait profiles. Unpublished doctoral thesis, University of Illinois, Urbana-Champaign, Ill., 1971. (*Dissertation Abstracts International,* 1971, *32,* (2-B), 1197.)

Stewart, T.R., Generality of multidimensional representations. *Multivariate Behavioral Research,* 1974, *9,* 507-519.

Stoop, I. Sekundaire analyse van de "Van jaar tot jaar data" met behulp van niet-lineaire multivariate technieken: Verschillen in de schoolloopbaan van meisjes en jongens. Research Bulletin, RB 001-80, Department of Data Theory, University of Leiden, Leiden, The Netherlands, 1980.

Svendrød, R. & Ursin, H. A factor analytic study of acquisition of a conditional emotional response in rats. *Journal of Comparative and Physiological Psychology,* 1974, *87,* 1174-1179.

Swaan, J. & Goossens, F.A. Handleiding bij de 'strange situation'. WEP-reeks, WR 82-24-EX, Vakgroep W.E.P., University of Leiden, Leiden, The Netherlands, 1982.

Takane, Y., Young, F.W., & De Leeuw, J. Non-metric individual differences multi-dimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika*, 1977, *42*, 7-67.

Ten Berge, J.M.F. *Optimizing factorial invariance*. (doctoral dissertation University of Groningen). Groningen, The Netherlands: Author, 1977.

Ten Berge, J.M.F. Comparing factors from different studies on the basis of factor scores, loadings, or weights. Technical Report, Department of Psychology, University of Groningen, Groningen, The Netherlands, 1982.

Teufel, S. TUCK, Tuckers Modell einer drei-dimensionalen Faktorenanalyse. Ein FORTRAN IV-Programm. In F. Gebhardt, *Statistische Programme des DRZ. Teil B: Einzelbeschreibungen*. Programm Information PI-33 des Deutschen Rechenzentrum, Darmstadt, 1969.

Thigpen, C.H. & Cleckley, H. A case of multiple personality. *Journal of Abnormal and Social Psychology*, 1954, *49*, 135-151.

Torgerson, W.S. *Theory and methods of scaling*. New York: Wiley, 1958.

Triandis, H.C. (Ed.) *The analysis of subjective culture*. New York, N.J.: Wiley-Interscience, 1972.

Triandis, H.C. (Ed.) *Variations in black and white perceptions of the social environment*. Urbana, IL: University of Illinois Press, 1976.

Triandis, H.C. Subjective culture and interpersonal relations across cultures. *Annals of the New York Academy of Sciences*, 1977, *285*, 418-434.

Triandis, H.C., Feldman, J.M., Weldon, D.E. & Harvey, W.M. Ecosystem distrust and the hard-to-employ. *Journal of Applied Psychology*, 1975, *60*, 44-56.

Triandis, H.C., Tucker, L.R., Koo, P. & Stewart, T. Three-mode factor analysis of the behavioral component of interpersonal attitudes. Technical Report No.50, Department of Psychology, University of Illinois, Urbana, IL, 1967.

Tucker, L.R. Determination of parameters of a functional relation by factor analysis. *Psychometrika*, 1958, *23*, 19-23.

Tucker, L.R. Implications of factor analysis of three-way matrices for measurement of change. In C.W. Harris (Ed.), *Problems in measuring change*. Madison: University of Wisconsin Press, 1963. (Pp.122-137).

Tucker, L.R. The extension of factor analysis to three-dimensional matrices. In H. Gullikson & N. Frederiksen (Eds.), *Contributions to mathematical psychology*. New York: Holt, Rinehart and Winston, 1964 (Pp. 110-119).

Tucker, L.R. Experiments in multimode factor analysis. In A. Anastasi (Ed.), *Testing Problems in Perspective*, Washington, DC: American Council on Education, 1966. Reprinted from: *Proceedings of the 1964 Invitational Conference on Testing Problems*. Princeton, N.J.: Educational Testing Service, 1965 (Pp. 46-57).

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

Tucker, L.R. Learning theory and multivariate experiment. Illustration by determination of generalized learning curves. In R.B. Cattell (Ed.), *Handbook of multivariate experimental psychology*. Chicago, IL: Rand McNally, 1966(b).

Tucker, L.R. Three-mode factor analysis of Parker-Fleishman complex tracking behavior data. *Multivariate Behavioral Research*, 1967, *2*, 139-151.

Tucker, L.R. Comments on "Confounding of sources of variation in factor-analytic techniques". *Psychological Bulletin*, 1968, *70*, 345-354.

Tucker, L.R. Relations between multidimensional scaling and three-mode factor analysis. *Psychometrika*, 1972, *37*, 3-27(a).

Tucker, L.R. Use of three-mode factor analysis in MDS. Paper presented at the Workshop on Multidimensional Scaling, University of Illinois, 7-10 June, 1972(b).

Tucker, L.R. Three-mode factor analysis applied to multidimensional scaling. Paper presented to the U.S.-Japan Seminar on Theory, Methods, and Applications of Multidimensional Scaling and Related Techniques, La Jolla, CA, August 20-24, 1975.

Tucker, L.R. & Messick, S. An individual differences model for multidimensional scaling. *Psychometrika*, 1963, *28*, 333-367.

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

Tukey, J.W. Discussion on the paper by Professor Cox and Mrs. Snell: "A general definition of residuals". *Journal of the Royal Statistical Society, Series B*, 1968, *30*, 248-275.

Tukey, J.W. *Exploratory data analysis*. Reading, Mass.: Addison-Wesley, 1977.

Tukey, J.W. & Wilk, M.B. Data analysis and statistics: An expository overview. *AFIPS Conference Proceedings, Fall Joint Comp(utational) Conference*, 1966, *29*, 695-709.

Tzeng, O.C.S. Differentiation of affective and denotative meaning systems in personality ratings via three-mode factor analysis. Unpublished doctoral thesis, University of Illinois, Urbana - Champaign, 1972. (*Dissertation Abstracts International*, 1973, *34*, (2-B), 864.)

Tzeng, O.C.S. Differentiation of affective and denotative meaning systems and their influence in personality ratings. *Journal of Personality and Social Psychology*, 1975, *32*, 978-988.

Tzeng, O.C.S. Differentiation of affective and denotative semantic subspaces. *Annals of the New York Academy of Sciences*, 1977, *285*, 476-500(a).

Tzeng, O.C.S. Individual differences in self-conception: multivariate approach. *Perceptual and Motor Skills*, 1977, *45*, 1119-1124(b).

Tzeng, O.C.S. & Landis, D. Three-mode multidimensional scaling with points of view solutions. *Multivariate Behavioral Research*, 1978, *13*, 181-213.

Van de Geer, J.P. Toepassing van drieweg-analyse voor de analyse van multiple tijdreeksen. In C.J. Lammers, Groei en ontwikkeling van de ziekenhuisorganisaties in Nederland. Interim rapport, Institute of Sociology, University of Leiden, Leiden, The Netherlands, 1974.

Van de Geer, J.P. Drieweg komponenten analyse (Memo). Department of Data Theory, University of Leiden, Leiden, The Netherlands, 1975.

Van de Geer, J.P. De analyse van arbeidscurven. *Nederlands Tijdschrift voor de Psychologie*, 1962, *17*, 233-246.

Van der Heijden, P. Het analyseren van verkiezingsuitslagen met behulp van correspondentieanalyse. *Kwantitatieve methoden*, 1982, *3*(8), 19-34.

Van der Kloot, W.A. & Kroonenberg, P.M., Group and individual implicit theories of personality: An application of three-mode principal component analysis. *Multivariate Behavioral Research*, 1982, *17*, 471-492.

Van der Kloot, W.A. & Van den Boogaard, T. Weights of information in impression formation. Report MT. 01-78, Psychological Institute, University of Leiden, Leiden, The Netherlands, 1978.

Van der Sanden, A.L.M. Het analyseren van structurele verandering: een methodologische aantekening ten behoeve van ontwikkelingspsychologie. Unpublished masters thesis, Department of Psychology, Catholic University of Nijmegen, Nijmegen, The Netherlands, 1973.

Van der Voort, T.H.A. *Kinderen en TV-geweld: waarneming en beleving*. Lisse, The Netherlands: Swets & Zeitlinger, 1982.

Van der Voort, T.H.A., Vooijs, M.W. & Bekker, P.A. TV-geweld in kinderogen. II Schaalconstructie, WR 82-06-EX, vakgroep W.E.P., University of Leiden, Leiden, The Netherlands, 1982.

Van Egeren, L.F., Headrick, M.W., & Hein, P.L. Individual differences in autonomic responses: Illustration of a possible solution. *Psychophysiology*, 1972, *9*, 626-633.

Van Maanen-Feijen, J.E. Analysis of multiple time-series. Research note, RN 008-68, Psychological Institute, University of Leiden, Leiden, The Netherlands, 1968.

Vavra, T.G. An application of three-mode factor analysis to product perception. In F.D. Allvine (Ed.), *Marketing in motion/ Relevance in Marketing*. Chicago: American Marketing Association, Series no.33, 1972 (Pp. 578-583)(a).

Vavra, T.G. Factor analysis of perceptual change. *Journal of Marketing Research*, 1972, *9*, 193-199(b).

Vavra, T.G. A three-mode factor analytic investigation into the effectiveness of advertising. Unpublished doctoral thesis, University of Illinois, Urbana-Champaign, Ill., 1973. (*Dissertation Abstracts International*, 1974, *34*, (12-A), 7802.)

Visser, R.A. *On quantitative longitudinal data in psychological research* (Doctoral thesis, University of Leiden). Leiden, The Netherlands: Author, 1982.

Wainer, H., Gruvaeus, G. & Blair, M. TREBIG: A 360/75 FORTRAN program for three-mode factor analysis designed for big data sets. *Behavioral Research Methods and Instrumentation*, 1974, *6*, 53-54.

Wainer, H., Gruvaeus, G. & Zill II, N. Senatorial decision making: I. Determination of structure. *Behavioral Science*, 1973, *18*, 7-19.

Walsh, J.A. An IBM 709 program for factor analyzing three-mode matrices. *Educational and Psychogical Measurement*, 1964, *24*, 669-773.

Walsh, J.A. & Walsh, R. A revised Fortran program for three-mode factor analysis. *Educational and Psychological Measurement*, 1976, *36*, 169-170.

Walter, J. Komplexe taaksituaties en hartsnelheidsvariabiliteit in de psychiatrie. Technical Report, Stichting Centrum St.-Bavo, Noordwijkerhout, The Netherlands, 1976.

Waters, E. The reliability and stability of individual differences in infant-mother attachment. *Child Development*, 1978, *49*, 483-494.

Wechsler, D. *Manual for the Wechsler Intelligence Scale for Children-Revised*. New York: The Psychological Corporation, 1974.

Weitzman, R.A. A factoranalytic method for investigating differences between groups of individual learning curves. *Psychometrika*, 1963, *28*, 69-80.

Wicker, F.W. A scaling study of synesthetic thinking. ETS Res. Bull. RB-66-25); doctoral thesis, Princeton, New Jersey, 1966. (*Dissertation Abstracts International,* 1966, *27*, (6-B), 2173).

Wicker, F.W. Mapping the intersensory regions of perceptual space. *American Journal of Psychology,* 1968, *81,* 178-188.

Wicker, F.W., Thorelli, I.M., Barron, III, W.L., & Ponder, M.R., Relationships among affective and cognitive factors in humor. *Journal of Research in Personality,* 1981, *15,* 359-370.

Wiggins, N. & Blackburn, M.C. Implicit theories of personality: an individual differences approach. *Multivariate Behavioral Research,* 1976, *11,* 267-285.

Wilk, M.B. & Gnanadesikan, R. Probability plotting methods for the analysis of data. *Biometrika*, 1968, *55*, 1-17.

Wishart, J. Growth-rate determinations in nutrition studies with the bacon pig and their analysis. *Biometrika*, 1938, *30*, 16-28.

Witzke, D.B. Determining developmental changes in Holtzman inkblot technique factors using three-mode factor analysis. Unpublished doctoral thesis, University of Texas, Austin, 1975. (*Dissertation Abstracts International,* 1975, *36*, (5-A), 2727.)

Wohlwill,J.F. *The study of behavioral development*. New York: Academic Press, 1973.

Wold, H. Path models with latent variables: The NIPALS approach. In H.M. Blalock, A. Aganbegian, F.M. Borodkin, R. Boudon, V. Capecchi (Eds.), *Quantitative sociology: International perspectives on mathematical and statistical modelling*. New York: Academic Press, 1975 (Pp.307-357).

Wolters, M. Dimensions of Dutch parties. Res. Bull. RB.003-75, Department of Data Theory, University of Leiden, Leiden, The Netherlands, 1975.

Young, F.W. Quantitative analysis of qualitative data. *Psychometrika*, 1981, *46*, 357-388.

Young, F.W., De Leeuw, J. & Takane, Y. Quantifying qualitative data. In E.D. Lantermann & H. Feger (Eds.), *Similarity and choice*. Wien, Austria: Hans Huber, 1980.

Young, F.W. & Lewyckyj, R. *ALSCAL-4 User's Guide*. Carrboro, N.C.: Data Analysis and Theory Associates, 1979.

Zenisek, T.J. Three-mode factor analysis via a modification of Tucker's computational method-III. *Educational and Psychological Measurement,* 1978, *38,* 787-792.

Zenisek, T.J., The measurement of job satisfaction: a three-mode factor analysis. Unpublished doctoral thesis, Ohio State University, 1980. (*Dissertation Abstracts International,* 1980, *41*, (1-A), 75).

Reference notes

1. Goossens, F.A. Working mothers and attachment (thesis in preparation).

2. Harshmann, R.A. & DeSarbo, W.S. Connotative congruence analysis: An application of PARAFAC to the selection of appropriate spokesmen for a given brand (manuscript in preparation).

3. Kroonenberg, P.M. & Van der Kloot, W.A. External three-mode principal component analysis (manuscript in preparation).

4. Van der Kloot, W.A., Bakker, D. & Kroonenberg, P.M. Implicit theories of personality: Further evidence of extreme response style (manuscript in preparation).

AUTHOR INDEX

SUBJECT INDEX

SAMENVATTING

Met dit boek worden drie doelen nagestreefd. In de eerste plaats is het een monografie over drieweg-hoofdassenanalyse, en er is gepoogd alle belangrijke aspecten van deze techniek te bespreken. Verreweg het grootste gedeelte van de literatuur op dit terrein, alsmede aanverwante gebieden uit de meerdimensionale schaalmethoden worden (kritisch) besproken.

In de tweede plaats worden verbeterde methoden gepresenteerd om drieweganalyse uit te voeren, alsmede de consequenties hiervan doorgelicht. Ook worden theoretische bijdragen gepresenteerd over het transformeren van kernmatrices naar wat 'simpele structuren' worden genoemd.

Het verschaffen van een handleiding van de methoden in de praktijk van het sociaal-wetenschappelijk onderzoek kan worden aangemerkt als een derde doel bij het schrijven van dit boek. Zowel in- als uitvoerproblemen en interpretaties worden in het algemeen besproken en in detail toegelicht bij de analyse van een dertiental voorbeelden. Om de bruikbaarheid voor de praktijk te verhogen is een (vrijwel compleet) overzicht gemaakt van toepassingen van drieweganalyse, die bovendien geclassificeerd zijn naar inhoudelijke onderwerpen, zodat een ieder aansluiting kan zoeken bij vakgenoten die eerder deze techniek hebben toegepast.

De eerste twee hoofdstukken hebben een algemeen karakter. Hoofdstuk 1 is een organisatorische wegwijzer voor het boek. Het bevat een leeswijzer, een woordenlijst en een overzicht van de gebruikte notatie. Hoofdstuk 2 is het boek zelf in een notedop en kan worden opgevat als een geannoteerde inhoudsopgave of als een handreiking voor diegenen die geen zin of tijd hebben om het hele boek door te lezen.

De hoofdstukken uit Deel I richten zich op de theoretische aspecten verbonden met drieweg-hoofdassenanalyse. In Hoofdstuk 3 worden de modellen (het Tucker3 en Tucker2 model) die de basis vormen voor de rest van het boek in de context geplaatst van andere modellen die op dit terrein zijn voorgesteld. Deze laatste modellen kunnen in twee klassen worden onderverdeeld, namelijk hoofdassenmodellen en factoranalyse-modellen, waarbij de laatste één stochastische weg hebben en bij de eerste klasse alle wegen niet-stochastisch zijn. De klasse van hoofdassen-modellen kan weer worden onderverdeeld in twee subklassen, namelijk die met drie 'ingedikte' wegen, zoals het Tucker3 model, drieweg-schaling, PARAFAC1 en INDSCAL, en die met twee ingedikte wegen zoals het Tucker2 model, PARAFAC2, IDIOSCAL, CANDECOMP en INDSCAL. Dit soort modellen wordt doorgaans geanalyseerd met behulp van alternerende kleinste-kwadratenmethoden, terwijl de factoranalyse-modellen worden geanalyseerd met technieken uit het domein van covariantie-structuurmodellen. Hoofdstuk 3 besluit met de bespreking van een aantal mogelijke toevoegingen aan de Tucker modellen, zoals schattingsprocedures over ontbrekende gegevens, optimale schaalprocedures voor gegevens van lagere meetniveaus, faciliteiten om externe analyses uit te voeren, uitbreidingen naar *n* wegen, etc.

Hoofdstuk 4 bevat de (technische) kern van het boek. Alternerende kleinste-kwadratenmethoden (ALS) voor de Tucker modellen worden gepresenteerd. Daarbij komen aan de orde: het bestaan van exacte en benaderende oplossingen, constructie van ALS-algoritmen, hun convergentie, alsmede een aantal andere technische details. Drie bestandjes worden gebruikt om de nauwkeurigheid en de correctheid van de algoritmen te bestuderen en om na te gaan hoe gevoelig de oplossingen zijn voor toenemende toevalsfluctuaties in de gegevens.

Deel I eindigt met een bespreking in Hoofdstuk 5 van methoden om kernmatrices zo te transformeren dat ze een eenvoudige structuur krijgen, bijvoorbeeld met veel nullen. Twee algoritmen om dit te bereiken worden besproken; de ene gebaseerd op orthonormale transformaties en de andere op niet-singuliere transformaties. De problemen wat betreft de laatste methode worden gesignaleerd. Twee bestanden worden gebruikt om de methoden te illustreren en te evalueren.

Ook Deel II bevat voornamelijk theoretische verhandelingen, maar nu ligt de nadruk op die problemen die voortkomen uit het toepassen van drieweg-hoofdassenanalyse in de praktijk. Drie gebieden worden benadrukt: het voorbewerken van de oorspronkelijke gegevens zodat ze geschikt zijn voor een drieweganalyse, het bewerken van de ruwe uitvoer zodat die gemakkelijk te interpreteren valt, en het analyseren van het gedeelte van de gegevens wat niet overeenkomt met het gebruikte drieweg-model.

Het eerste deel van Hoofdstuk 6 geeft een overzicht van voorstellen die gedaan zijn om ruwe gegevens zo te bewerken dat ze geschikt zijn voor een (drieweg-)hoofdassenanalyse. Met name methoden om verschillen in gemiddelden en varianties te elimineren worden behandeld; hierbij wordt een verschil gemaakt tussen (on) interpreteerbare en (on)vergelijkbare gemiddelden en varianties. Een aantal modellen die gebruik maken van interpreteerbare gemiddelden en varianties worden besproken, met name die welke bestaan uit additieve termen voor de gemiddelden, multiplicatieve termen voor de varianties en product-termen voor de componenten. Ook wordt het probleem van iteratieve standaardizatie aangestipt.

In het tweede deel van Hoofdstuk 6 wordt er aandacht geschonken aan het interpreteren van de resultaten van een drieweganalyse en aan methoden om deze interpretatie te vergemakkelijken en te verbeteren. Hierbij passeren het schalen en interpreteren van componenten en de kernmatrix de revue, alsmede het gebruik van 'gezamenlijke componentenruimtem' en hun afbeeldingen, en verder het gebruik van component scores in drieweganalyse.

In Hoofdstuk 7 wordt de functie en het nut van residuenanalyse besproken voor (drieweg-)hoofdassenanalyse. Ook worden procedures en gedetailleerde aanbevelingen voor het behandelen van drieweg-residuen gepresenteerd en geïllustreerd. Met name komen aan bod: variantieanalyse op gekwadrateerde residuen, kwadratensommengrafieken, en het gebruik van normale kansverdelingsgrafieken.

In Deel II wordt de theorie uit de voorgaande delen toegepast en toegelicht met behulp van diverse bestanden. Elk van de geanalyseerde bestanden is opgenomen als een vertegenwoordiger van een bepaalde klasse toepassingen; onderzoekers met soortgelijke gegevens kunnen mogelijk hieruit inspiratie putten voor een geschikte aanpak voor hun eigen analyse.

Hoofdstuk 8 *(Attachment study)* bevat een gedetailleerde analyse van de reactie van kleine kinderen op een gestandaardiseerde procedure die de aard en de mate van gehechtheid van die kinderen aan hun moeder poogt te meten. De gegevens, scores van individuen op een aantal variabelen onder verschillende condities, zijn kenmerkend voor het soort gegevens dat met vrucht via drieweganalyse kan worden onderzocht.

Hoofdstuk 9 *(Triple personality study)* bevat een voorbeeld van gegevens verzameld met behulp van semantische differentialen, waarbij met name de individuele verschillen van groot belang zijn. Verschillen in gebruik van de relaties tussen schalen en begrippen staan centraal in deze analyse. De gegevens zijn afkomstig van één vrouw met drie verschillende persoonlijkheden: Eve Black, Eve White en Jane. De behandeling van de gegevens kan als voorbeeld dienen voor andere onderzoeken waarbij vergelijkbare beoordelingsschalen of testen worden gebruikt.

Hoofdstuk 10 *(ITP study)* bevat een voorbeeld van wat asymmetrische gelijkenisgegevens zou kunnen worden genoemd. In plaats van een gedetailleerde behandeling van de gegevens wordt dit hoofdstuk gebruikt om de 'theoretische proefpersoon' te introduceren, dat wil zeggen een proefpersoon geconstrueerd op basis van theoretische inzichten. Zo'n persoon is met name nuttig om als leidraad te gebruiken bij de interpretatie van de hoofdassen in een proefpersoonsruimte.

Hoofdstuk 11 *(Cola study)* bevat een heranalyse van gegevens over de smaak van cola's, waarbij met name aandacht wordt besteed aan de vergelijking van gelijkenisoordelen en oordelen verzameld met beoordelingsschalen. De resultaten van de drieweganalyse op de gelijkenisoordelen worden vergeleken met die van een drietal schaalmethoden voor individuele verschillen, aangezien dit soort gegevens doorgaans met deze methoden wordt geanalyseerd.

Hoofdstuk 12 *(Four ability-factor study)* laat zien hoe drieweganalyse kan worden gebruikt om correlatiematrices te analyseren. Hiermee wordt een mogelijkheid aangegeven om transversale gegevens eventueel uit verschillende bronnen te onderzoeken. Vooral in die gevallen waarin slechts de correlatiematrices gepubliceerd zijn, maar niet de oorspronkelijke gegevens kan een dergelijke analyse nuttig zijn. Het voorbeeld behandelt de analyse van matrices met correlaties tussen verschillende intelligentietesten, die afkomstig zijn van normale en zwakzinnige kinderen van verschillende leeftijden.

Hoofdstuk 13 *(Hospital study)* behandelt de analyse van multivariate longitudinale gegevens met behulp van drieweganalyse, geïllustreerd met een voorbeeld uit de organisatiesociologie. In de studie worden de structuren in de organisatie van Nederlandse ziekenhuizen bekeken over de jaren heen. Verschillende problemen met betrekking tot de analyse van dergelijke gegevens worden besproken, met name de relatie tussen drieweganalyse en autoregressie-modellen.

Hoofdstuk 14 *(Learning-to-read study)* bevat een uitbreiding van Tucker's werk over leercurven. De leercurven in deze studie zijn afkomstig van kinderen die in de loop van een jaar leren lezen en ze worden gevormd door de scores op een aantal testen die hun leesvordering meten. Een vluchtige vergelijking met resultaten van een lineaire logistische analyse op dezelfde gegevens wordt gemaakt.

Hoofdstuk 15 *(Leiden electorate study)* bevat een wat afwijkend voorbeeld, aangezien het gaat om de analyse van residuen uit kruistabellen met frequenties, in plaats van een analyse van gemeten gegevens. De uitslagen van Leidse kiesdistricten in drie verschillende verkiezingen vormen de basis voor de analyse. Allereerst worden de relevante interacties in de gegevens bepaald met loglineaire analyse; vervolgens zijn de residuen van een aantal loglineaire modellen onderzocht met behulp van drieweganalyse. De gebruikte procedure is een drieweg-generalisatie van correspondentieanalyse.

Errata

Errata

"Three-mode Principal Component Analysis: Theory and Applications"

(including a very selective list of new three-mode papers)

Pieter M. Kroonenberg

DSWO Press, Wassenaarseweg 52, Leiden, The Netherlands
P. O. Box 9555, 2300 RB Leiden

Page	Line	Correction
10	12ff	These lines should read: 'but it specifies common angles between the axes of the stimulus space. However, differential weighting of these axes is allowed.'
18	6	Formula should read: $h_{j1}s_{i11} + h_{j2}s_{i21}$.
19		In Figure 2.1 The labels 'Standard PCA' and 'Q-PCA' should be interchanged.
25	8	'SS(Data)' should read 'SS(Total)'.
25	10	'SS(Data$_f$)' should read 'SS(Total$_f$)'.
25	-12	Delete two sentences: 'Similarly [..] information'.
26	15	'j-centring' should read 'j-centring /standardization'.
26	19	'jk-centring' should read 'jk-centring /standardization'.
48	- 1	'183, 186' should read '83,86'.
48ff		Chapter 3. Kiers (in press-a, in press-b) has presented two new taxonomies for three-way models. His treatment includes several models from the French school (see also Carlier, et al., 1989, and Lavit, 1988).
49		Figure 3.1. The statement for going from IDIOSCAL to PARAFAC2 is incorrect. The proper expression can be found on p. 55.
49	-5	model~~s~~
50	2-4	The 'number of parameters' indicated in the table are not corrected for dependence between the parameters. e.g. for the Tucker3 model, the number

		of independnet parameters should be decreased by '$s^2+t^2+u^2$', and the Tucker2 model by 's^2+t^2'.
50	4	In the table stubs 'diff.' and 'red.' should be interchanged.
50	7	Three-mode scaling: '21s' should read '1s'.
51	10-13	The Tucker2 and Tucker3 models are not necessarily orthonormal, in fact Tucker did nor require this. The orthonormality is an expedient constraint for solving the estimation problem, and may be dropped later, as can for instance be seen in Chapter 5.
53	3	'$c_{pp'r} = c_{pp'r}$' should read '$c_{pp'r} = c_{p'pr}$'.
53	-10	'p.90-92, and section 6.2)' should read 'p.90-92), and section 6.2'.
56	18,19	The 'weak conditions' mentioned here are too strongly formulated, see the references mentioned.
57	13	In contrast to what is implied in the text, the expression $Z = GC(H'\blacksquare E')$ refers to a 'lateral plane' representation, i.e. both Z and C are juxtaposed lateral planes. The proper expression for a 'frontal plane' representation as used verbally in the text and in the next formula is $Z = GC(E'\blacksquare H')$.
58	4	'W =' should read 'W_k ='.
58	12,13	Such procedures have been recently discussed by Harshman, Kruskal, and Lundy in various permutations (see reference list).
61	3	'for the communalities' should read 'for factor analysis by estimating the communalities'.
62	8	'estimate' should read 'estimating'.
64	- 4	S and s should be underlined.
65	4	S ans s should be underlined.
68		Section 3.7. One of the reviewers missed a section on resampling plans, such as the bootstrap (Efron, 1982). One application using this approach is contained in Kroonenberg & Snyder (1989).
70	-10,-12	The papers referred to in this section have now been published: Van der Kloot & Kroonenberg (1985) and Van der Kloot, Kroonenberg, & Bakker (1985).
80	- 2	'c_{pqr}' should read '$\hat{c}_{pqr}$ '

80	20	'Penrose (1955 -' should read 'Penrose (1956 -', and the reference given on p. 376 is incorrect (see below).
82		It has been shown in Brouwer (1985) that the SS(Fit)/SS(Total) = R^2, given the data have been centred such that the overall mean is zero.
83	5	Formula should start with a minus sign.
83	- 6	Index of last summation sign should be 'r' rather than " q' ".
83	11	The statement 'the Hessian is negative' is not applicable for the present maximization with restrictions. Correct approaches may, for instance, be found in Luenberger (1973, p. 226).
84	15	'only the assess' should read 'only to assess'.
84	- 2	Add " $z_{i'j'k'}$ " to the end of the formula.
86		Two alternatives have been proposed for the TUCKALS algorithm. Weesie & Van Houwelingen (1983) constructed an algorithm based solely on regression techniques, rather than on eigenvalue-eigenvector decompositions, by including the estimation of the core matrix into the iterative process. The advantage of their approach is that missing data can be handled in a natural straightforward way. Murakami (1983) produced an ALS algorithm for the Tucker2 model which uses the multivariable-multioccasion matrix as its starting point. This gives the possibility of analysing published matrices of this kind, and is an alterantive for the Invariant Factors Model of McDonald (1984).
87-89		Recently Kroonenberg, Ten Berge, Brouwer, & Kiers (1989) have shown that the Bauer-Rutishauser step in the TUCKALS algorithms may be replaced by the modified Gram-Schmidt orthogonalization procedure. The latter is slightly faster than the former.
103		The proof assumes orthonormality, while at p.81, it was suggested that the proof should be valid without this assumption. A correct proof is given in Ten Berge, De Leeuw, & Kroonenberg (1987).

105		A different more direct proof is contained in Ten Berge, De Leeuw, & Kroonenberg (1987).
110		The discussion on this page should have referred to Green (1952).
111	12	Delete from (5.3): ' and $K'K = I_s$ '
113		The diagonality problem (NS) is essentially equivalent with Harshman's PARAFAC (Harshman, & Lundy, 1984a). A more detailed report on this equivalence is contained in Brouwer (1985), and related material is given by Harshman, Kruskal, and Lundy in various permutations (see reference list).
128		Sections 6.1-6.8. The issue of centring and standardization has recently received a much more detailed and algebraic treatment (Harshman & Lundy, 1984b; Kruskal, 1984). Some additional commentary see Harshman and Lundy (1985).
131		'pca-data' are commonly called 'profile data'; 'mds-data' are commonly called 'proximity data'; 'anova-data' resemble 'conjoint measurement data'. see Shepard (1972).
138ff		Three-mode data. The ANOVA-first approach (see also p.195) has been treated in considerable detail by Kettenring (1983a,b) using the PARAFAC model for the three-way decomposition. An application with the Tucker models can be found in Kroonenberg & Van der Voort (1987).
141	- 5	'$g_{ip}h_{jq}e_{kr}$' should be '$g_{ip}h_{jp}e_{kp}$'.
151	3-9	In contrast to the statement, averages of correlations <u>are</u> correlations.
153	5	Before the third summation sign an '=' should be inserted.
154	14ff	Components as latent elements. Probably all that can be said is that the major components span a space which captures most of the common variability. Whether there are directions in this space which correspond to latent entities or theoretical constructs is a different matter, and thus the statements in this section stand to be corrected.

155	8	'session' should read 'section'.
155ff		Rotation of components. Most reviewers of the book have chided the author for not treating this issue in more depth. (see the reviews for detailed arguments).
159	18,19	The diagonal matrix lambda should be absorbed into H rather than G (see also remark, p. 19).
160	1-4	A diagonal cork matrix as suggested here, can only occur if $\Sigma_{p,q}\ c_{pqr}c_{pqr'} = 0$, as required by the all-orthogonality of the core matrix (see Weesie & Van Houwelingen, 1983). Thus $G_1 = \begin{matrix} 2 & 0 \\ 0 & 2 \end{matrix}$ and $G_2 = \begin{matrix} -1 & 0 \\ 0 & 1 \end{matrix}$ is possible, but $G_1 = \begin{matrix} 2 & 0 \\ 0 & 2 \end{matrix}$ and $G_2 = \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}$ is not. The cited example indeed shows the required pattern.
160	7-15	The diagonal/antidiagonal phenomenon is also in concordance with the all-orthogonality.
164	- 3	'$\widetilde{HV}_r$' should read 'HV_r'.
165	2,3	'When C_r [..] be used' should read 'When C_r is not square with, say s<t, C_r selects as-dimensional subspace from the t-dimensional space of H'.
167	-4,5	'$h_{.q}$' should read '$\bar{h}_{.q}$'.
180	20	'1.67' should read '1.64'.
202ff		Chapter 8: The final and corrected data have now been reported in Van IJzendoorn, et al. (1985) with, however, very condensed information on the three-mode analysis.
227ff.		Chapter 9. Triple personality data. A reworked version of this chapter was published as Kroonenberg (1985).
257		Section 11.2. From the introduction of this section one might get the impression that unlike methods for multidimensional scaling three-mode principal component analysis needs similarities. The remarks made refer to the fact that the output of the three-mode programs are easiest to read when high values indicate closeness, rather than separateness. Such

an effect can easily be obtained by converting dissimilarities into inner products by the standard Torgerson procedure for classical multidimensional scaling. When the input data themselves are already considered proportional to inner products, then high values should indicate closeness.

274ff Chapter 12: The reference to Levin (1966) was omitted in this chapter. An extensive treatment of the analysis of sets of correlation/covariance matrices is now available in Kroonenberg & Ten Berge (1987, 1989). Furthermore, insufficient justice was done to Procrustes procedures (see Gower, 1975; Ten Berge, 1982).

Chapter 12: The analysis is in this chapter is performed on a set of correlation matrices. There ae several reasons why such an analysis is less than desirable (Harshman and Lundy, 1984, p.141). Especially within the context of structural equation modeling, there is strong opposition to analysing correlation rather than covariance matrices (see e.g. Jöreskog, 1971; McDonald, 1984, 3p.292; Meredith, 1984). The main argument centres around the different size of one standard deviation unit for the same variable across different samples.

275 7ff The second paragraph on p. 275 contains somewhat confusing statements about eigenvalues. The following remarks taken from Harshman's review clarify the issue: "When performing a two-way analysis using a singular value decomposition (SVD) of raw data, the eigenvalues and related sums of squares are obtained from the **squared** singular values, whereas when doing an SVD of a covariance or correlation matrix, one examines the **unsquared** singular values to obtain the eigenvalues and related sums of squares of the original data from which the covariances were computed. Likewise, when TMFA is applied to raw data, one looks at squared core elements, but when it is applied to

		covariances, one should examine the unsquared core elements. If this is done, "eigenvalues" obtained from two-way and three-way methods are directly comparable." (p.331).
285ff.		Chapter 13. For a different analysis of the hospital data see Kroonenberg, Lammers, & Stoop (1985).
313ff.		Chapter 14. These data were also analysed by curve fitting with logistic regression, see Jansen & Bus (1984).
322		The column means in Table 14.4 should read: '12 13 17 27' giving column effects '-5 -4 0 10', and residuals for Q of '-4 -2 1 0 2 0'. Eliminating the 'maybe too many positive scores'of Q (1. -4,-5), and invalidating the remark about Q.
326ff		Chapter 15. Strictly speaking, it is not correct to use the name 'correspondence analysis' in this chapter, as the basic properties of correspondence analysis do not hold. For a further investigation into a proper three-mode correspondence analysis, see Kroonenberg (1989).
327	5	Delete '+ log r_{ij}'
339	- 1	'-18' should read '-8'.
360	19	'Nesselroode' should read 'Nesselroade'.
362	16	'model of application' should read 'model by application'.
364		Insert after Einhorn: Fienberg, S.E. **The analysis of cross-classified categorical data** (2nd edition). Cambridge, MA: MIT Press, 1980.
371	- 0	'1987' should read '1978'.
376	6	'791' should read '591'.
376	16-18	The correct reference is: Penrose, R. (1956). On the best approximate solutions of linear matrix equations. **Proceedings of the Cambridge Philosophical Society**, 52, 17-19.
377	- 9	'alternative' should read 'alternating'.
393		Entry Joint plots: '163-165' should read '164-166'.

References

Brouwer, P. (1985). **The TUCKALS-PARAFAC connection.** Unpublished doctoral thesis, Vakgroep Algemene Pedagogiek, Department of Education, Leiden University.

Carlier, A., Lavit, Ch., Pagès, M., Pernin, M.O., & Turlot, J.C. (1989). In R. Coppi, & S. Bolasco (Eds.), **Analysis of multiway data matrices.** Amsterdam: Elsevier.

Efron, B. (1982). **The jackknife, the bootstrap, and other resampling plans.** Philadelphia: Society for Industrial and Applied Mathematics.

Gower,J.C. (1975). Generalized Procrustes analysis. **Psychometrika, 40,** 33-51.

Green, B.F. (1952). The orthonormal approximation of an oblique structure in factor analysis. **Psychometrika, 17,** 429-440.

Harshman, R.A., & Lundy, M.E. (1984a). The PARAFAC model for three-way factor analysis and multidimensional scaling. In H.G. Law, C.W. Snyder Jr, J.A. Hattie, & R.P. McDonald (Eds.), **Research methods for multimode data analysis** (pp.122-215). New York: Praeger.

Harshman, R.A., & Lundy, M.E. (1984b). Data preprocessing and the extended PARAFAC model. In H.G. Law, C.W. Snyder Jr, J.A. Hattie, & R.P. McDonald (Eds.), **Research methods for multimode data analysis** (pp. 216-284). New York: Praeger.

Harshman, R.A., & Lundy, M.E. (1985). **The preprocessing controversy: An exchange of papers between Kroonenberg, Harshman and Lundy.** (Research Bulletin No. 631). London, Ontario: University of Western Ontario, Department of Psychology. (Includes a contribution of Kroonenberg (1984), Report No. WR 84-54-IN).

Harshman, R.A., Lundy, M.E., & Kruskal, J.B. (1989). Comparison of trilinear and quadrilinear methods: Strengths, weaknesses, and degeneracies. Internal Report. Department of Psychology. University of Western Ontario, London, Canada.

Jansen, M.G.H., & Bus, A.G. (1984). Individual growth patterns in early reading performance. **Kwantitatieve Methoden,** 5(14), 97-109.

Jöreskog, K. G. (1981). Simultaneous factor analysis in several populations. **Psychometrika, 36,** 409-426.

Kettenring, J.R. (1983a). Components of interaction in analysis of variance models with no replications. In P.K. Sen (Ed.), **Contributions to statistics: Essays in honor of Norman L. Johnson** (pp.283-297). Amsterdam: North Holland .

Kettenring, J.R. (1983b). A case study in data analysis. **Proceedings of symposia in Applied Mathematics, 28,** 105-139.

Kiers, H.A.L. (in press). Hierarchical relations among three-way methods. **Psychometrika.**

Kiers, H.A.L. (1988). Comparison of "Anglo-Saxon" and "French" three-mode methods. **Statistique et Analyse de Données, 13,** 14-32.

Kroonenberg, P.M. (1985). Three-mode analysis of semantic differential data: The case of a triple personality. **Applied Psychological Measurement, 9,** 83-94.

Kroonenberg, P.M. (1989). Singular value decomposition of interactions in three-way contingency tables. In R. Coppi, & S. Bolasco (Eds.), **Analysis of multiway data matrices** (pp. 169-184). Amsterdam: Elsevier.

Kroonenberg, P.M., Lammers, & Stoop, I. (1985). Three-mode principal component analysis of multivariate longitudinal organizational data. **Sociological Methods and Research, 14,** 99-136.

Kroonenberg, P.M., & Snyder Jr., C.W. (1989). Individual differences in assimilation resistance and affective responses in problem solving. **Multivariate Behavioral Research, 24.**

Kroonenberg. P.M., & Ten Berge, J.M.F. (1987). Cross-validation of the WISC-R factorial structure using three-mode principal component analysis and perfect congruence analysis. **Applied Psychological Measurement, 11,** 195-210.

Kroonenberg, P.M., Ten Berge, J.M.F., Brouwer, P., & Kiers, H.A.L. (1989). Gram-Schmidt versus Bauer-Rutishauser in alternating least-squares algorithms for three-way data. **Computational Statistics Quarterly.**

Kroonenberg, P.M., & Ten Berge, J.M.F. (1989). Three-mode principal component analysis and perfect congruence analysis for sets of covariance matrices. **British Journal of Mathematical and Statistical Psychology, 42,** 63-80.

Kroonenberg, P.M., & Van der Voort, T.H.A. (1987). Multipiicatieve decompositie van interacties bij oordelen over de werkelijkheidswaarde van televisiefilms [Multiplicative decomposition of interactions for judgments of realism in television films]. **Kwantitatieve Methoden, 8**(23), 117-144.

Kruskal, J.B. (1984). Multilinear methods. In H.G. Law, C.W. Snyder Jr, J.A. Hattie, & R.P. McDonald (Eds.), **Research methods for multimode data analysis** (pp. 36-62). New York: Praeger.

Kruskal, J.B., Harshman, R.A., & Lundy, M.E. (1989). Some relationships among Tucker 3-mode factor analysis (3-MFA), PARAFAC, and CANDELINC. In R. Coppi, & S. Bolasco (Eds.), **Analysis of multiway data matrices** (pp. 115-122). Amsterdam: Elsevier.

Levin, J. (1966). Simultaneous factor analysis of several Gramian matrices. **Psychometrika, 31,** 413-419.

Luenenberger, D.G. (1973). **Introduction to linear and nonlinear programming**. Reading, MA: Addison-Wesley.

Lundy, M.E., Harshman, R.A., & Kruskal, J.B. (1989). A two-stage procdure incorporating good features of both trilinear and quadrilinear methods. In R. Coppi, & S. Bolasco (Eds.), **Analysis of multiway data matrices** (pp. 123-130). Amsterdam: Elsevier.

McDonald, R.P. (1984). The invariant factors model for multimode data. In H.G. Law, C.W. Snyder Jr, J.A. Hattie, & R.P. McDonald (Eds.), **Research methods for multimode data analysis** (pp. 285-307). New York: Praeger.

Meredith, W. (1964). Notes on factorial invariance. **Psychometrika, 29,** 177-185.

Murakami, (1983). Quasi three-mode principal component analysis - A method for assessing the factor change. **Behaviormetrika, 14,** 27-4$\overset{4}{8}$

Penrose, R. (1956). On the best approximate solutions of linear matrix equations. **Proceedings of the Cambridge Philosophical Society, 52,** 17-19.

Shepard, R.,N. (1972). A taxonomy of some principal types of data and multidimensional methods for their analysis. In R.N.

Shepard, A.K. Romney, & S.B. Nerlove (Eds.), **Multidimensional scaling: theory and applications in the behavioral sciences** (Vol. I, Theory) (pp.23-47). New York: Seminar Press.

Ten Berge, J.M.F. (1982). Orthogonal Procrustes rotation for two or more matrices. **Psychometrika,42**, 267-276.

Ten Berge, J.M.F., De Leeuw, J., & Kroonenberg, P.M. (1987). Some additional results on principal components analysis of three-mode data by means of alternating least squares algorithms. **Psychometrika, 52**, 183-191.

Van der Kloot, W.A., & Kroonenberg, P.M. (1985). External analysis for three-mode principal component models. **Psychometrika, 50**, 479-494.

Van der Kloot, W.A., Kroonenberg, P.M., & Bakker, D. (1985). Implicit theories of personality: Further evidence of extreme response style. **Multivariate Behavioral Research, 20**, 369-387.

Van IJzendoorn, M.H., Goossens, F.A., Kroonenberg, P.M., & Tavecchio, L.W.C. (1985). Dependent attachment: B4 children in the Strange Situation. **Psychological Reports, 57**, 439-451.

Weesie, J. & Van Houwelingen, J. (1983). **GEPCAM Users' manual. Generalized principal components analysis with missing data.** Institute of Mathematical Statistics, University of Utrecht.

Recent books on three-way data analysis

Arabie, P, Carroll, J.D., & De Sarbo, W.S. (1987). **Three-way scaling and clustering.** Sage University Paper Series on Quantitive Applications in the Social Sciences, Series No. 65. Beverly Hills: Sage.

Coppi, R., & Bolasco, S. (Eds.) (1989). **Analysis of multiway data matrices.** Amsterdam: Elsevier.

Flury, B. (1989). **Common principal component analyis and related multivariate models.** New York: Wiley.

Lavit, C. (1988). **Analyses conjointe de plusieurs matrices des données.** Paris: Masson.

Law, H.G., Snyder Jr, C.W., Hattie, J.A.,& McDonald, R.P. (1984). **Research methods for multimode data analysis.** New York: Praeger.

Reviews of the book

Everitt, B.S. (1986). **Biometrics, 42,** 224-225.

Finkbeiner, C.T. (1984). **Journal of Classification, 1,** 133-138.

Harshman, R.A. (1984). **Applied Psychological Measurement, 9,** 327-332.

Koch, W.G. (1984). **Contemporary Psychology, 29,** 915-916.

Partzsch, U. (1988). **Biometrical Journal, 30,** 482.

Ten Berge, J.M.F. (1985). **Kwantitatieve Methoden, 6**(15), 145-147, (in Dutch).

Summary of the book

Dissertation Abstracts International - Section C, 1984, 45, 501.

Three-mode principal component analysis
Theory and applications
Pieter M. Kroonenberg

In multivariate analysis the data are usually contained in a single matrix with n rows and m columns, corresponding with n individuals and m variables. Or, to put it differently, the data have two modes: individuals and variables. It has already been known for a long time that this particular two-mode representation of data is too restrictive in a number of very important cases. Often there are three modes, the additional mode being replications, occasions, conditions, points-of-view, and so on. The data must be collected in a three-mode matrix, which has n rows, m columns, and k slices. Of course this 'data-box' can be flattened into an ordinary two-way matrix in various ways, but often there is no unique obvious way in which this should be done. Moreover most data analysis techniques that can be applied to the two-way matrix obtained by summation or concatenation over one of the modes, simply ignore the fact that the data were originally three mode.

In the early sixties Tucker introduced a form of principal component analysis which works directly on the three-mode matrix, and has parameters for all three modes. This was an enormous step ahead, because there was no need to flatten databoxes any more. Somewhat later individual differences models were introduced in multidimensional scaling. They are also based on three-mode data, and pretty soon the two developments were integrated by Tucker and Carroll. For the individual differences models specialized algorithms are available, but for Tuckers's three-mode component model the available algorithms were somewhat ad hoc and suboptimal.

In **Three-mode principal component analysis** perhaps the main emphasis is on the development and study of a satisfactory algorithm for Tuckers's technique, together with a friendly computer program. But this is not all. Models for three-mode data are also discussed in considerable detail. Important data analytic decisions, which must be taken before the programs can be applied, are spelled out. A chapter on the analysis of residuals shows that the job is not done if the program has run. About half of the book is used to analyze meticulously a number of examples, which are chosen in such a way that each of them is representative for a large class of data structures. There are semantic differentials, three-way contingency tables, replicated correlation matrices, three-way similarity data, and multivariate time-series data. All examples are used to show which plots can be made, how the residuals should be analyzed, how the key parameters should be interpreted, and so on.

Three-mode principal component analysis says about everything there is to say about this class of techniques. The book does this on a mathematical and computational level, but more importantly it illustrates everything that is said by using a number of very real and very interesting examples.

DSWO PRESS

ISBN 90 6695 002 1