

SCALING AND INTERPRETATION

6

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

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

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) \times variables (j-mode) \times 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) \times variables or stimuli (j-mode) \times 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 ($\ell \times m \times n$) 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 ($\mu, \beta, \gamma, \zeta$), with standar-

dization 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:

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

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. $g_{i1} = 1/\sqrt{l}$ 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

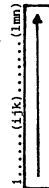
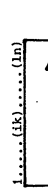

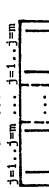
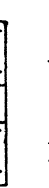
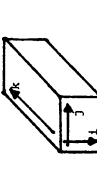
$$\begin{aligned}
 z_{ijk} = & c_{111}/\sqrt{\ell mn} + (1/\sqrt{\ell m}) \sum_{r=2}^u e_{kr} c_{11r} + (1/\sqrt{\ell n}) \sum_{q=2}^t h_{jq} c_{1q1} + \\
 & (1/\sqrt{mn}) \sum_{p=2}^s g_{ip} c_{p11} + (1/\sqrt{\ell}) \sum_{q=2}^t \sum_{r=2}^u h_{jq} e_{kr} c_{lqr} + \\
 & (1/\sqrt{m}) \sum_{p=2}^s \sum_{r=2}^u g_{ip} e_{kr} c_{p1r} + (1/\sqrt{n}) \sum_{p=2}^s \sum_{q=2}^t g_{ip} h_{jq} c_{pq1} + \\
 & \sum_{p=2}^s \sum_{q=2}^t \sum_{r=2}^u g_{ip} h_{jq} e_{kr} c_{pqr} .
 \end{aligned}$$

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:

$$\begin{aligned}
 z_{ijk} = & c_{111} + e_{k1} + h_{j1} + g_{i1} + \sum_{p=2}^s h_{jp} e_{kp} + \sum_{p=2}^s g_{ip} e_{kp} + \\
 & \sum_{p=2}^s g_{ip} h_{jp} + \sum_{p=2}^s g_{ip} h_{jq} e_{kr},
 \end{aligned}$$

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 ⁺	one-way means to be retained	for selected possibility	eliminated formula means	data arrangement [†] for centring	references [§]
1. none		all	none	$\bar{z}_{ijk} = z_{ijk}$		Tucker (1966)-a; V. d. Geer (Note 4)-1; TUCKALS-option 0.
2. over all data points z_{ijk}	overall centring	all	$z_{...}$	$\bar{z}_{ijk} = z_{ijk} - z_{...}$		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)	$z_{.j.}$	$\bar{z}_{ijk} = z_{ijk} - z_{.j.}$		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 (absolute) jk-centring (positive) ki-centring (normative)	between subjects (i)	$z_{.jk}$	$\bar{z}_{ijk} = z_{ijk} - z_{.jk}$		jk-centring Tucker-o; Bartussek-a; 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 modes separately; double-centring	ij, jk-centring (performative) jk, ki-centring (performative) ki, ij-centring	between conditions (k)	$z_{.jk} z_{.ik}$	$\bar{z}_{ijk} = z_{ijk} - z_{.jk} - z_{.ik} + z_{.k}$		general: Gollob (1968b, c); Tucker (1968). jk, ki-centring TUCKALS-option 1.
6. triple-centring		none	$z_{.jk} z_{.ik} z_{.il}$	$\bar{z}_{ijk} = z_{ijk} - z_{.jk} - z_{.ik} - z_{.il} + z_{.j.} + z_{.i.} + z_{.l.} - z_{...}$		

⁺ 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)-z means that this way of centring was labelled a by Tucker in his 1966 paper.

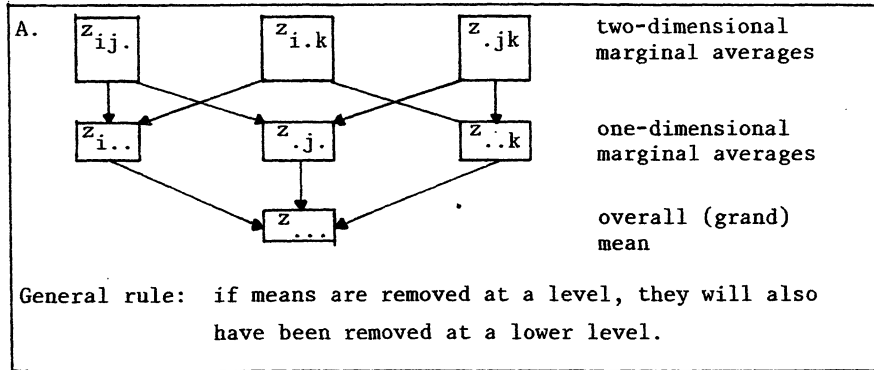
6.5 INPUT CENTRING: TYPES, CONSEQUENCES, RECOMMENDATIONS

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



B.

case type means removed consequences
type of centring

2	$z_{...}$ <i>overall centring</i>	no components centred
3	$z_{.j.}$ <i>j-centring</i>	ik-combination-mode components centred (e.g. in Tucker's Method III) component scores on j-mode components centred
4	$z_{.jk}$ <i>jk-centring</i>	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 $tu \times tu$ latent covariation matrix becomes covariance matrix
5	$z_{.jk}, z_{i.k}$ <i>jk, ik-centring</i>	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 $c(tu \times tu)$ latent covariance matrix T2: (j-mode unreduced): i-mode components centred c_{pjr} centred per pr-combination $(su \times su)$ latent covariance matrix

6	$z_{.jk}, z_{i.k}, z_{ij}$	T3: i-mode, j-mode, k-mode components centred
	<i>triple-centring</i>	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:

$$\begin{aligned} \tilde{z}_{ijk} - \tilde{z}_{.jk} &= \sum_{p,q,r} g_{ip} h_{jq} e_{kr} c_{pqr} - (1/\ell) \sum_i \sum_{p,q,r} g_{ip} h_{jq} e_{kr} c_{pqr} = \\ &= \sum_{p,q,r} h_{jq} e_{kr} c_{pqr} \{g_{ip} - (1/\ell) \sum_i g_{ip}\} = \\ &= \sum_{p,q,r} h_{jq} e_{kr} c_{pqr} \{g_{ip} - \bar{g}_{.p}\}, \end{aligned}$$

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 centring (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

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

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

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

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.

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

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' out-

put 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

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

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

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

are the estimated data values based on the fitted model, and

$$\sum_{p=1}^s \lambda_p^{\times} = \sum_{q=1}^t \mu_q^{\times} = \sum_{r=1}^u v_r^{\times} = SS(\text{Fit})$$

with λ_p^{\times} , μ_q^{\times} , and v_r^{\times} the eigenvalues.

- Standardized eigenvalues or standardized component weights, λ_p ($p=1, \dots, s$), μ_q ($q=1, \dots, t$), v_r ($r=1, \dots, u$), are the eigenvalues divided by the total variation in the data expressed by

$$SS(\text{Total}) = \sum_{i=1}^{\ell} \sum_{j=1}^m \sum_{k=1}^n z_{ijk}^2.$$

- Core matrices are scaled such that

$$\sum_{p=1}^s \sum_{q=1}^t \sum_{r=1}^u c_{pqr}^2 = SS(\text{Fit}) \quad (\text{T3}), \text{ or}$$

$$\sum_{k=1}^n \sum_{p=1}^s \sum_{q=1}^t \tilde{c}_{pqk}^2 = SS(\text{Fit}) \quad (\text{T2});$$

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 analogously 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	λ_p^x μ_q^x ν_r^x	$\lambda_p^* = \lambda_p^x / mn$ $\mu_q^* = \mu_q^x / \ell n$ $\nu_r^* = \nu_r^x / \ell m$
components	g_{ip} h_{jq} e_{kr}	$g_{ip}^* = g_{ip} \sqrt{\lambda_p^*} = g_{ip} \sqrt{(\lambda_p^x / mn)}$ $h_{jq}^* = h_{jq} \sqrt{\mu_q^*} = h_{jq} \sqrt{(\mu_q^x / \ell n)}$ $e_{kr}^* = e_{kr} \sqrt{\nu_r^*} = e_{kr} \sqrt{(\nu_r^x / \ell m)}$
core matrix	c_{pqr}	$c_{pqr}^* = c_{pqr} / \sqrt{(\lambda_p^* \mu_q^* \nu_r^*)} = \ell mn \times c_{pqr} / \sqrt{\lambda_p^x \mu_q^x \nu_r^x}$

Note: g_p , h_q , and e_r are orthonormal components;
 λ_p^x , μ_q^x , and ν_r^x 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 interpreted. 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

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

three modes, and c_{221} the strength of the relationship between the second components of the first and the second mode in combination

with the first of the third mode. The interpretation of the elements of the core matrix is facilitated if one realizes that

$$\sum_{p=1}^s \sum_{q=1}^t \sum_{r=1}^u c_{pqr}^2 = \text{SS}(\text{Fit}) \text{ (T3)}, \text{ and } \sum_{p=1}^s \sum_{q=1}^t \sum_{k=1}^n \tilde{c}_{pqk}^2 = \text{SS}(\text{Fit}) \text{ (T2)}$$

In other words, each of the c_{pqr}^2 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 analogous interpretation holds for the T2 core matrix, but now a \tilde{c}_{pqk}^2 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}_{pqk}^2 as proportion $\frac{\tilde{c}_{pqk}^2}{\sum_p \sum_q \tilde{c}_{pqk}^2}$, in order to indicate the importance of that particular combination for the k-th condition or subject.

In addition it can be shown that

$$\sum_{p=1}^s \sum_{q=1}^t c_{pqr}^2 / \text{SS}(\text{Total}) = v_r \quad (r=1, \dots, u)$$

$$\sum_{q=1}^t \sum_{r=1}^u c_{pqr}^2 / \text{SS}(\text{Total}) = \lambda_p \quad (p=1, \dots, s)$$

$$\sum_{r=1}^u \sum_{p=1}^s c_{pqr}^2 / \text{SS}(\text{Total}) = \mu_q \quad (q=1, \dots, t)$$

with λ_p , μ_q , and v_r the standardized component weights for the first, second, and third mode respectively. Similar expressions can be derived for the T2 core matrix. Furthermore

$$\begin{aligned} \sum_{p=1}^s \sum_{q=1}^t \sum_{r=1}^u c_{pqr}^2 / \text{SS}(\text{Total}) &= \sum_{p=1}^s \lambda_p = \sum_{q=1}^t \mu_q = \sum_{r=1}^u v_r \\ &= \text{SS}(\text{Fit}) / \text{SS}(\text{Tot}) = \text{Rel. SS}(\text{Fit}) \end{aligned}$$

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, $Z = GAH'$, 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, $G\Lambda$ (see also section 2.2, Figure 2.1). The strength of the interaction is usually expressed as λ_{pp}^2 , 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 ($s \times s \times s$) 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 $s \times s \times u$ core matrix a still relatively 'simple structure' is $c_{pqr} \neq 0$ if $p=q$, $r=1, \dots, u$, 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 $r=2$:

$$c_{pqr} \neq 0 \text{ for } p=q, r=1; c_{pqr} \neq 0 \text{ for } q=s-p+1, r=2;$$

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 help-

ful. 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 compo-

nent 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 $\tilde{c}_{ppk}^{\frac{1}{2}}$ and $\tilde{c}_{qqk}^{\frac{1}{2}}$, 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 ($p=1, \dots, s$) and h_q ($q=1, \dots, t$) - 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 .

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:

$$D_r = G C_r H' = G(U_r \wedge_r V_r')H' = \left(\frac{\ell}{m}\right)^{\frac{1}{2}} (G U_r \wedge_r)^{\frac{1}{2}} \left(\frac{m}{\ell}\right)^{\frac{1}{2}} (H V_r \wedge_r)^{\frac{1}{2}} = \tilde{G}_r \tilde{H}_r'$$

with

$$\tilde{G}_r = \left(\frac{\ell}{m}\right)^{\frac{1}{2}} G U_r \wedge_r^{\frac{1}{2}} \quad \text{and} \quad \tilde{H}_r = \left(\frac{m}{\ell}\right)^{\frac{1}{2}} H V_r \wedge_r^{\frac{1}{2}} \quad r=1, \dots, u.$$

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_{ij}^r , are the scalar product between the vectors, i.e.

$$d_{ij}^r = \sum_{p=1}^a \tilde{g}_{ip} \tilde{h}_{jp}, \text{ with } a = \min(s,t).$$

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 suc-

cesful 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

$$d_{ijr} = \sum_{p=1}^s \sum_{q=1}^t g_{ip} h_{jq} c_{pqr} \text{ or } D_r = G C_r H',$$

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

$$\sum_{p=1}^a \tilde{g}_{ip}^r \tilde{h}_{jp}^r,$$

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 $M(1,2) = \{m_{iq}^{12}\}$, or $\{m_{iq}\}$ for short, is defined as

$$m_{iq} = \sum_{k=1}^n \sum_{j=1}^m \sum_{p=1}^s \sum_{r=1}^u g_{ip} h_{jq} e_{kr} c_{pqr} = \sum_{k=1}^n \sum_{j=1}^m h_{jq} d_{iqk},$$

$$\text{with } d_{iqk} = \sum_{p=1}^s \sum_{r=1}^u g_{ip} e_{kr} c_{pqr}.$$

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

$$\begin{aligned} m_{iq} &= n \sum_{j=1}^m h_{jq} \left\{ \frac{1}{n} \sum_{k=1}^n d_{iqk} \right\} = n \sum_{j=1}^m h_{jq} \bar{d}_{iq} = n \bar{d}_{iq} \left\{ \sum_{j=1}^m h_{jq} \right\} = \\ &= n \bar{d}_{iq} \bar{h}_{.q}. \end{aligned}$$

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 $\bar{h}_{.q}$. For centred data entire mixed-mode matrices may become zero, because all h_{jq} , and/or analogous 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.

