PART I

THREE-WAY METHODS APPLIED TO QUANTIFICATION MATRICES

2. A HIERARCHY OF THREE-WAY METHODS

Many methods have been developed for the analysis of three–way data. In the present chapter, a number of these will be discussed. The methods discussed here form a hierarchy. The methods in the hierarchy are related such that while going down the hierarchy one finds a method that represents the data by a simpler model, albeit at the cost of a poorer representation of the variables. It should be noted that this hierarchy is an extension of a similar hierarchy mentioned by Kroonenberg (1983, pp.49 ff). A different hierarchy of three–way methods was described by Carroll and Wish (1974, pp.92–96). They describe hierarchical relations between IDIOSCAL, PARAFAC2, and INDSCAL.

The three–way methods to be discussed here are all methods that can be applied to quantification matrices. Quantification matrices will be described in detail in chapter 3. For the purpose of the present chapter only some notational aspects of the quantification matrices need to be mentioned. Let m be the number of variables, n be the number of objects, and S_j, $j = 1,...,m$, be the $n \times n$ quantification matrix for variable j. As will be seen in chapter 3, quantification matrices can often be considered as similarity matrices between the objects, hence the choice S_j for the symbol to denote such matrices. It should be noted that the S_j matrices, being similarity matrices, are always symmetric. The data to be handled by the three–way methods described below consist of an $n \times n \times m$ array, that is to say, of a set of m matrices of order $n \times n$. Figure 2.1 depicts such a data array.

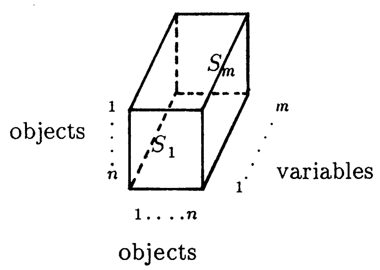

Figure 2.1. A three–way data array for similarity matrices.

In the next four sections methods for analyzing such data will be described. In section 2.5 it will be shown how these methods form a hierarchy.

2.1. STATIS

STATIS has been developed by L'Hermier des Plantes (1976) as a method for performing PCA on a set of quantification matrices in three steps. The first step, called STATIS–1 here, consists of performing PCA on the matrices $S_1,...,S_m$ considered as variables. That is, STATIS–1 searches linear combinations, $F_1,...,F_r$, of the matrices $S_1,...,S_m$ that optimally account for the matrices $S_1,...,S_m$. In order to see how the matrices $S_1,...,S_m$ can be considered as variables, let matrix S_j be represented by a vector $\text{Vec}(S_j)$ which contains the elements of S_j strung out row-wise, $j = 1,...,m$. Similarly, let $\text{Vec}(F_1),...,\text{Vec}(F_r)$ be the principal components of the variables $\text{Vec}(S_1),...,\text{Vec}(S_m)$. Then, STATIS–1 can be described as the method that minimizes the loss function

$$\text{STATIS–1}(F_1,...,F_r,C) = \sum_{j=1}^{m} \| S_j - \sum_{l=1}^{r} c_{jl} F_l \|^2$$

$$= \sum_{j=1}^{m} \| \text{Vec}(S_j) - \sum_{l=1}^{r} c_{jl} \text{Vec}(F_l) \|^2 \qquad (1)$$

over arbitrary matrices $F_1,...,F_r$, and the $(m \times r)$ matrix C of loadings c_{jl} of the variables on the components. Collecting the variables $\text{Vec}(S_1),...,\text{Vec}(S_m)$ in the $n^2 \times m$ data matrix S, and the components $\text{Vec}(F_1),...,\text{Vec}(F_r)$ in the $n^2 \times r$ matrix F, the STATIS–1 function can be rewritten as

$$\text{PCA}(F,C) = \| S - FC' \|^2. \qquad (2)$$

This description of the loss function for STATIS–1 is simply a description of the loss function for PCA (r components) on a data set S of order $n^2 \times m$, where F contains the component scores for the n^2 object–pairs, and C contains the component loadings for the variables. This method has been mentioned earlier by Tucker (1972, pp.7–8) in developing his three-mode scaling method.

The second step of STATIS consists of defining a compromise matrix as the first principal component (F_1) of the matrices $S_1,...,S_m$. That is, assuming that α_j gives the first principal component weight for matrix S_j, $j = 1,...,m$, then this compromise matrix is given by $F_1 = \sum_j \alpha_j S_j$.

The third step, which is called STATIS–3 here, consists of PCA of the compromise matrix that has been defined in the second step. It is readily verified that PCA on matrix F_1, and thus STATIS–3, is equivalent to minimizing

$$\text{STATIS–3}(X,\Lambda) = \| \sum_{j=1}^{m} \alpha_j S_j - X\Lambda X' \|^2 \qquad (3)$$

over the diagonal matrix Λ, and matrix X ($n \times r$) subject to $X'X = I_r$. At the minimum of the STATIS–3 function, matrix X will contain the compromise component scores for the objects, and Λ will contain the corresponding eigenvalues.

The strategy of analyzing a number of matrices as if they are variables followed by a detailed analysis of the summarizing matrices has been proposed earlier by Tucker and Messick (1963) in a related context. Tucker and Messick have proposed to analyze the lower triangles of a number of *distance* matrices by means of PCA on these lower triangles strung out as vectors of order $\frac{1}{2}n(n-1)$. Typically, these components are rotated to simple structure. Next, new "distance matrices" that optimally approximate the original distance matrices are computed on the basis of each of these principal components, which are subsequently analyzed by means of classical multidimensional scaling techniques. Although the Tucker and Messick method resembles STATIS in lay–out, it clearly differs from STATIS in several respects.

2.2. TUCKALS–3

Tucker (1966) proposed various models for three–mode principal components analysis. One of these is the model which Kroonenberg and De Leeuw (1980) called the Tucker–3 model. This model represents the entries of each of the three modes by means of a smaller number of components, hence providing three sets of components. These components are related to each other by the so–called core matrix. In case this model is applied to symmetric matrices $S_1,...,S_m$ it can be described as

$$\hat{s}_{ii'j} = \sum_{u=1}^{p} \sum_{v=1}^{p} \sum_{l=1}^{r} x_{iu} x_{i'v} c_{jl} h_{uvl} \ , \qquad (4)$$

where $\hat{s}_{ii'j}$ is the model–description of element (i,i') of matrix S_j, x_{iu} denotes the coordinate of object i on component u, c_{jl} the loading of variable

j on component l, and h_{uvl} the element of the core matrix that relates the u^{th} and v^{th} components for the objects to the l^{th} component for the variables, $u, v = 1,...,p$, and $l = 1,...,r$.

Fitting the symmetric version of the Tucker–3 model (4) in the least squares sense (TUCKALS–3) comes down to minimizing the function

$$\text{TUCKALS-3}(X,H_1,...,H_r,C) = \sum_{j=1}^{m} \| S_j - X \sum_{l=1}^{r} c_{jl} H_l X' \|^2, \qquad (5)$$

over matrices X $(n \times p)$, $H_1,...,H_r$ $(p \times p)$, and C $(m \times r)$. Kroonenberg (1983) has called this method three–mode scaling. According to Kiers (1989b) this symmetric version of TUCKALS–3 can be described as a constrained variant of PCA on the matrices $S_1,...,S_m$. In order to show this, we rearrange the elements of the matrix between $\|\ \ \|$ into vectors, and use the fact that $\text{Vec}(XH_lX') = (X \otimes X)\text{Vec}(H_l)$, where $\otimes$ denotes the Kronecker product. Then the loss function for TUCKALS–3 can be rewritten as

$$\begin{aligned}\text{TUCKALS-3}(X,H_1,...,H_r,C) &= \sum_{j=1}^{m} \| \text{Vec}(S_j) - \text{Vec}(X \sum_{l=1}^{r} c_{jl} H_l X') \|^2 \\ &= \sum_{j=1}^{m} \| \text{Vec}(S_j) - \sum_{l=1}^{r} (X \otimes X)\text{Vec}(H_l) c_{jl} \|^2 \\ &= \| S - (X \otimes X)(\text{Vec}\, H_1 | \ldots | \text{Vec}\, H_r) C' \|^2. \qquad (6)\end{aligned}$$

Minimizing (6) over arbitrary matrices X $(n \times p)$, $H_1,...,H_r$ $(p \times p)$, and C $(m \times r)$ is equivalent to minimizing the PCA loss function (2) over C and F, subject to the constraint that F $(n^2 \times r)$ can be written as $F = (X \otimes X)(\text{Vec}\, H_1 | \ldots | \text{Vec}\, H_r)$ for certain matrices X and $H_1,...,H_r$ of appropriate orders. Thus, TUCKALS–3 can be seen as constrained PCA, where matrix C yields loadings for the variables.

Whereas STATIS–1 only gives a representation of the variables, TUCKALS–3 also gives a representation of the objects (in X). In addition to these object coordinates TUCKALS–3 provides measures that indicate the interaction–relations between different components for the objects and the variables (given in the matrices $H_1,...,H_r$). The latter relations, however, are difficult to interpret, because they are relations between objects components as "viewed" by each of the components for the variables (indicated by the subscripts of the matrices $H_1,..,H_r$).

2.3. INDSCAL and INDORT

Carroll and Chang (1970) have proposed INDSCAL for analyzing a set of distance or dissimilarity matrices. Their first step is transforming the distances into similarities by means of the Torgerson (1958) transformation. Next they propose to fit the similarities to the INDSCAL model, given by

$$\hat{s}_{ii'j} = \sum_{l=1}^{r} x_{il} x_{i'l} c_{jl} , \tag{7}$$

where x_{il} denotes the object coordinate of object i on component l, and c_{jl} gives the loading of variable j on component l, $l = 1,...,r$. This model is much simpler than the Tucker-3 model. The interpretational difficulties in TUCKALS-3, concerning the matrices $H_1,...,H_r$, are overcome by INDSCAL. The INDSCAL model does not consider relations between object components and variable components. In this model only one set of r components is defined. These components can be interpreted as components for objects and variables simultaneously, which makes interpretation of the results much easier than interpreting the results of a TUCKALS-3 analysis. Interpreting the TUCKALS-3 results is further complicated by the fact that the solutions of TUCKALS-3 have rotational freedom. The INDSCAL model, on the other hand, does not allow for rotation of its components. It provides unique axes.

Fitting the INDSCAL model in the least squares sense comes down to minimizing

$$\text{INDSCAL}(X,W_1,...,W_m) = \sum_{j=1}^{m} \| S_j - XW_jX' \|^2, \tag{8}$$

over an $n \times r$ matrix X of object coordinates and diagonal matrices $W_1,...,W_m$. In order to describe INDSCAL in terms of PCA of the $n^2 \times m$ matrix S, the elements of the matrix between $\|\ \ \|$ are again rearranged in vectors. Now it is useful to note that $\text{Vec}(XW_jX') = \text{Vec}(\sum_l \mathbf{x}_l w_{jl} \mathbf{x}_l') = \sum_l \text{Vec}(\mathbf{x}_l \mathbf{x}_l') w_{jl}$, where $\mathbf{x}_l$ is the l^{th} column of X, and w_{jl} is the l^{th} diagonal element of W_j. Let $c_{jl} \equiv w_{jl}$, then INDSCAL can be described in terms of PCA on the $n^2 \times m$ data matrix S as minimizing

$$\text{INDSCAL}(X,C) = \sum_{j=1}^{m} \| \text{Vec}(S_j) - \text{Vec}(XW_jX') \|^2$$

$$= \sum_{j=1}^{m} \| \text{Vec}(S_j) - \sum_{l=1}^{r} \text{Vec}(\mathbf{x}_l\mathbf{x}_l')c_{jl} \|^2$$

$$= \| S - (\text{Vec}(\mathbf{x}_1\mathbf{x}_1')|\ldots|\text{Vec}(\mathbf{x}_r\mathbf{x}_r'))C' \|^2. \quad (9)$$

Minimizing (9) over arbitrary X and C is equivalent to minimizing the PCA loss function (2) over C and F, subject to the constraint that F can be written as $F = (\text{Vec}(\mathbf{x}_1\mathbf{x}_1')|\ldots|\text{Vec}(\mathbf{x}_r\mathbf{x}_r'))$ for some $n \times r$ matrix X.

In INDSCAL the constraints imposed on PCA are stronger than those for TUCKALS–3, provided that the number of components for the objects (p) is larger than or equal to the number of components for the variables (r). This can be seen by verifying that INDSCAL can be considered as TUCKALS–3 with the matrices $H_1,\ldots,H_r$ constrained such that $h_{uvl} = 1$ when $u=v=l$, and 0 otherwise, provided that $p \geq r$, (cf. Kroonenberg, 1983, p.53, where a thus constrained core matrix is called the "three–way analogue of an identity matrix"). Of course, the advantages of the stronger and simpler INDSCAL model are offset by the expected loss of fit of this more heavily constrained version of PCA.

Instead of simply minimizing (9) over arbitrary matrices X, one may minimize (9) subject to the constraint $X'X = I_r$. Kroonenberg (1983, p.118) denotes this method as "orthonormal INDSCAL". Here, it will be denoted by the acronym INDORT. Being a constrained variant of INDSCAL, INDORT is a constrained variant of PCA. This method is of special interest in the present study.

2.4. SUMPCA

Levin (1966) has developed a method for the simultaneous factor analysis of a number of data sets. His method is based on PCA of the sum of a set of matrices, $S_1,\ldots,S_m$. His method is equivalent to one of the stages in Tucker's three mode Principal Components Analysis (Tucker, 1966). As has been shown by Jaffrennou (1978), this in turn is equivalent to one of the stages of Jaffrennou's method for analyzing a three–way array. Finally, Gower (1966) proposed to analyze a dissimilarity matrix by first applying the Torgerson transformation in order to obtain similarities, and next finding coordinates for the objects by means of PCA on the similarity matrix. The latter similarity matrix is often computed as the sum of a number of similarity matrices expressing the similarities between the same objects in terms of different variables. Then this method, "Principal Coordinates Analysis", can be seen as equivalent to the other three methods mentioned above. Because this method comes down to performing a PCA on the sum of the similarity matrices it is called "SUMPCA" here.

SUMPCA can be described mathematically as minimizing the function

$$\text{SUMPCA}(X,\Lambda) = \| \sum_{j=1}^{m} S_j - X\Lambda X' \|^2$$

$$= m \sum_{j=1}^{m} \| S_j - X(m^{-1}\Lambda)X' \|^2 + \text{constant}, \qquad (10)$$

over X $(n \times r)$, subject to $X'X = I_r$, and over the diagonal matrix Λ.

From the description of SUMPCA as the method that minimizes (10) it is clear that STATIS–3, see (3), is a weighted variant of SUMPCA. When all weights $\alpha_1,...,\alpha_m$ in STATIS–3 are (taken) equal, then SUMPCA and STATIS–3 coincide. Alternatively, STATIS–3 can be seen as the SUMPCA method applied to quantification matrices $\alpha_j S_j$ instead of S_j.

SUMPCA can be described as a constrained variant of PCA as follows. SUMPCA minimizes

$$\text{SUMPCA}^*(X,W) = \sum_{j=1}^{m} \| S_j - XWX' \|^2 \qquad (11)$$

over X and W, where $W = m^{-1}\Lambda$, subject to the constraint that X is column–wise orthonormal, and W is diagonal. Defining $c_{jl} \equiv w_l$, hence requiring that c_{jl} be the same for all j, and making the same derivation as for (9), we have

$$\text{SUMPCA}^*(X,W) = \| S - (\text{Vec}(\mathbf{x}_1\mathbf{x}_1')|...|\text{Vec}(\mathbf{x}_r\mathbf{x}_r'))C' \|^2. \qquad (12)$$

SUMPCA is a constrained variant of PCA, in that it minimizes (2) over F subject to the constraint that F can be written as $F = (\text{Vec}(\mathbf{x}_1\mathbf{x}_1')|...|\text{Vec}(\mathbf{x}_r\mathbf{x}_r'))$ for a certain column–wise orthonormal matrix X, and over C subject to the constraint that all rows of C are equal. SUMPCA is a constrained variant of PCA that is even more heavily constrained than INDORT, because of the additional constraint imposed on C. This constraint implies that the components do not weight the variables differentially, contrary to INDORT and INDSCAL.

2.5. Hierarchical relations between three-way methods

As discussed above and in Kiers (1989b), all methods described here are constrained versions of STATIS–1. The methods have been treated in such an order that each method is a constrained version of its predecessor. That is, in reversed order, SUMPCA is a constrained version of INDORT, INDORT is a constrained version of INDSCAL, INDSCAL is a constrained version of TUCKALS–3 (provided that in TUCKALS–3 the number of components for the objects, p, is not smaller than r), and TUCKALS–3 is a constrained version of STATIS–1.

In Table 2.1 an overview is given of the hierarchy formed by these methods. Each method can be seen as a method that fits a model to the data. The data are represented by the $n^2 \times m$ matrix S, the model prediction by $\hat{S}$. In addition, the constraints imposed on the model parameters are given. In Table 2.1 two new symbols are introduced, H for $H \equiv (\text{Vec } H_1 | \ldots | \text{Vec } H_r)$, and $\mathbf{c}$ for the r–vector with elements w_l, $l = 1,\ldots,r$, for describing $C = \mathbf{1c}'$, with $\mathbf{1}$ an m–vector with unit elements.

Table 2.1. *A hierarchy of three – way methods.*

method	model	additional constraints
STATIS–1	$\hat{S} = FC'$	
TUCKALS–3	$\hat{S} = (X \otimes X)HC'$	
INDSCAL	$\hat{S} = (\text{Vec}(\mathbf{x}_1\mathbf{x}_1') \| \ldots \| \text{Vec}(\mathbf{x}_r\mathbf{x}_r'))C'$	
INDORT	$\hat{S} = (\text{Vec}(\mathbf{x}_1\mathbf{x}_1') \| \ldots \| \text{Vec}(\mathbf{x}_r\mathbf{x}_r'))C'$	$X'X = I_r$
SUMPCA	$\hat{S} = (\text{Vec}(\mathbf{x}_1\mathbf{x}_1') \| \ldots \| \text{Vec}(\mathbf{x}_r\mathbf{x}_r'))C'$	$X'X = I_r$, $C = \mathbf{1c}'$

As has been mentioned above, TUCKALS–3 only fits in the hierarchy, when $p \geq r$, because only in that case TUCKALS–3 provides a better fit for $S_1, \ldots, S_m$ than INDSCAL does. TUCKALS–3 models with $p < r$ cannot be located in this hierarchy.

The hierarchy given in Table 2.1 holds for any choice of quantification matrix. In chapter 4 an overview is given of these methods for different choices of quantification matrices for qualitative variables and for mixtures of qualitative and quantitative variables, respectively.

2.6. Suggestions for an eclectic approach to three-way analysis of a set of quantification matrices

Above, it has been shown that a number of well-known three-way methods can be ordered in a hierarchy. The higher the position a method takes in this hierarchy, the better the representation of the variables is. Simultaneously, the higher the position a method takes in the hierarchy, the more parameters are involved in the model, and hence the more complex the model is. The latter statement may not be obvious for the methods that take the highest positions in the hierarchies. That is, it may seem that STATIS–1 uses less parameters than TUCKALS–3. However, STATIS–1 in fact fits the "full" Tucker–3 model, that is, with $p = n$, as is readily verified. Thus, STATIS–1 does fit a model with more parameters than TUCKALS–3.

A larger number of parameters in a model typically makes interpretation of the results more complex. For example, TUCKALS–3 provides coordinates for the objects which are related to the variables in a complicated way by means of an extra set of "core" parameters. The coordinates for the objects provided by INDSCAL are linked in a simpler way to the components for the variables, because every component for the variables refers to exactly one component for the objects. In SUMPCA the components for the variables are trivialized in that all variables have the same loadings on each dimension. SUMPCA fits the simplest model, because the model is in fact no longer a three-way model: it does not give a differential representation of the variables.

In order to perform a PCA of the variables, one might choose from all of the methods above. No general statement as to which method is the *best* can be made. However, the hierarchy described above might be used in order to find empirically which method is the most *useful* for describing one's data by means of a PCA of the variables. Obviously, the *best* representation of the variables is provided by STATIS–1. However, for the purpose of interpretation of the solution this method is rather poor, because it does not yield any description (coordinates) for the objects that is linked to the principal components for the variables. At the other extreme, SUMPCA only yields a good description of the relations between the variables when these are strong. A useful strategy might be to start analyzing one's data by means of the method at the bottom of the hierarchy, that is by SUMPCA. If SUMPCA yields a sufficient fit and an interpretable solution in a reasonable number of dimensions it gives the simplest possible representation of the data (in that number of dimensions). If this method does not give an adequate solution one may start "climbing" the hierarchy and analyze the data by means of the next method in the hierarchy (with the same number of dimensions). This procedure can be repeated until one finds a solution that adequately represents the variables and the objects, if such a solution is available at all. Of course, decisions about "adequate" representations, or "reasonable" numbers of dimensions will always be based on subjective evaluation to some extent.

3. THE CHOICE OF QUANTIFICATION MATRICES

3.1. Why use quantification matrices?

In the previous chapter a series of methods has been discussed for the analysis of three-way data. In chapter 1 it has been mentioned that these three-way methods can all be applied to a set of quantification matrices. However, it has not yet been explained what a quantification matrix is, nor why it should be used. These questions will be dealt with in the present section. First, the possible types of variables, for which these quantification matrices are to be defined, will be reviewed.

Variables can be distinguished according to their levels of measurement. Although many refinements in such a distinction are possible the following distinction will be made here. Variables are called "qualitative" (or "nominal") if the "scores" of the observation units (objects) on such variables do not contain any numerical information at all. Examples of such variables are a person's nationality, a person's religion, an animal's genus, a vegetable's taste, etc.

Variables are called "ordinal" if the scores of objects on the variables have a predefined ordering, possibly with ties. An object that falls in a higher category can be said to have "more" of the aspect that is measured by the variable. For example, the rank order score after a world championship soccer, or a person's preference listing of a number of food-items can be considered as ordinal variables.

Finally, variables are called "quantitative" (or "numerical" or "interval level") when the scores on the variables have a numerical meaning. That is, scores on quantitative variables do not only indicate the rank order of the objects, but also how much the objects differ in the aspect measured by the variable. A person's length, an object's volume, a tree's number of leaves, etc, are usually considered to be quantitative variables.

The above distinction in three types of variables may seem very strict. In practice, however, it is often not at all clear whether a variable can be considered to be measured at interval level or at ordinal level. Similar ambiguities may arise among qualitative and ordinal variables. Therefore, it is important to note that the level of measurement of a variable is not a given property, but a property attributed to the variable by the practitioner. It is this person who decides at what level of measurement a variable is considered to be measured. This choice is not only based on what kind of variables one has under study, but also on what aspect of the variables one wants to analyze. It can be useful, for instance, to consider variables as "length" and "weight" as ordinal if one wants to detect nonlinear relations between such variables.

The main purpose of the present study is to describe techniques for the analysis of (mixtures of) variables that are considered qualitative or quantitative variables. As has been explained above, qualitative and quantitative variables are of a very different kind. One cannot compare scores on a qualitative variable with those on a quantitative variable, for instance, because these scores have completely different meanings. This implies that one cannot calculate (ordinary) correlation coefficients between variables of different measurement levels. It would be useful to have a means of comparing such different variables. One way of doing so is by representing each variable (qualitative or quantitative) by means of what is called a "quantification matrix" (Zegers, 1986, p.26). Here, a quantification matrix is a square matrix containing measures of similarity among the objects in terms of the variable at hand. Because these quantification matrices are of the same order and of the same kind (being similarity matrices between objects), one can compare quantification matrices for different variables, regardless whether or not they are considered at the same level of measurement. It will be explained later why such matrices can be seen as similarity matrices. First, however, it will be described for what purposes quantification matrices have been proposed and how they have been used.

The idea of using quantification matrices emerged from the wish to define "correlation coefficients" for variables of a measurement level lower than the interval level. The term "correlation coefficient" is used here in a slightly wider sense than usual. That is, the term is used for any type of association coefficient that can be seen as a scalar product between two normalized sets of scores, which do not necessarily have to be deviation scores. The definition of correlation measures for quantification matrices that represent the variables has been developed independently by different authors. Daniels

(1944) has used square matrices of order n to represent ordinal variables, and showed that for certain choices of these matrices their normalized inner products give Spearman's and Kendall's rank correlation coefficients, respectively.

Another line of research based on quantification matrices finds its origin in the work of Escoufier (1970, 1973). He proposed the notion of correlation coefficients (RV–coefficients) between so–called "operators", which are square matrices containing scalar products between certain sets of (scores) vectors. The correlation between such operators is computed simply as the normalized scalar product between the vectors containing all elements of the operators in some fixed order.

Saporta (1975) used this notion of correlation coefficients for operators (quantification matrices) in order to find correlation coefficients for qualitative variables. A qualitative variable can be considered a set of indicator variables. Each indicator variable indicates whether an object falls in a category (score 1) or not (score 0). These indicator variables are collected in an indicator matrix. For variable j this indicator matrix is denoted as G_j, of order $n \times m_j$, and $D_j \equiv G_j'G_j$, the diagonal matrix of order m_j with category frequencies on the diagonal, where n is the number of objects and m_j is the number of categories of variable j. Saporta (1975) chooses as a quantification matrix for a qualitative variable, the matrix $JG_jD_j^{-1}G_j'J$, where $J = (I - n^{-1}\mathbf{1}\mathbf{1}')$ is the centering operator, and $\mathbf{1}$ is the n–vector with unit elements. The correlation coefficient defined as the correlation between such quantification matrices, that is, $\mathrm{tr}\,S_j'S_l/(\mathrm{tr}\,S_j^2)^{½}(\mathrm{tr}\,S_l^2)^{½}$ with $S_j = JG_jD_j^{-1}G_j'J$, and $S_l = JG_lD_l^{-1}G_l'J$ for variables j and l, can be seen as a correlation between qualitative variables.

Vegelius (1973) followed the same strategy for defining correlations between variables of low measurement level. He proposed to define correlation coefficients for variables of low measurement level as correlations between quantification matrices, that is, again $\mathrm{tr}\,S_j'S_l/(\mathrm{tr}\,S_j^2)^{½}(\mathrm{tr}\,S_l^2)^{½}$, and called such correlation coefficients "E–coefficients", because they represent scalar products in Euclidean space. Together with Janson, Vegelius considered various correlation coefficients for (mixtures of) qualitative, ordinal and quantitative variables (e.g., Janson & Vegelius, 1978a, 1978b, 1982). The construction of correlation coefficients on the basis of quantification matrices has been reviewed by Marcotorchino (1984) and Zegers (1986).

Although the use of quantification matrices was inspired by the wish to define correlation coefficients between variables of low measurement level, in this study quantification matrices are not used for determining correlations between such pairs of variables, but for techniques that simultaneously analyze a set of qualitative and quantitative variables. The idea of simultaneously analyzing a set of quantification matrices has been given by Escoufier (1973) as well as by Vegelius (1973), who both propose to analyze the matrix of correlation coefficients between quantification matrices by means of Principal Components Analysis (PCA). However, if one is not satisfied with an analysis of the variables only, but wants to have a representation of the objects as well, it is necessary to analyze these quantification matrices by means of other methods. This has motivated Cazes, Bonnefous, Baumerder and Pagès (1976) to extend Saporta's method. As is explained in section 5.1, their method does not fully succeed in simultaneously analyzing both the variables and the objects. Later on, D'Ambra and Marchetti (1986), see also Coppi (1986), have suggested to use other approaches to analyze a set of quantification matrices. In chapter 4 methods based on this suggestion are described.

Above, it has been explained why one might use quantification matrices in order to analyze (mixtures of) qualitative and quantitative variables. In the next sections, a number of possible quantification matrices will be discussed, both for qualitative and for quantitative variables.

3.2. Quantification matrices for qualitative variables

The quantification matrices, denoted by S_j, to be proposed in the present section can all be described in terms of the notation given above. A full account of all quantification matrices that have been proposed in the literature is beyond the scope of this study. The following summary of quantification matrices (Table 3.1) is based mainly on the correlation coefficients for nominal variables mentioned by Zegers (1986, pp. 50–53). Some quantification matrices have not been given explicitly in the form as they are described in Table 3.1, but only implicitly by means of the correlation coefficients that are based on them. Therefore, in addition to the quantification matrix itself, the corresponding correlation coefficients are given, if available.

Table 3.1. *Quantification matrices for a qualitative variable.*

quantification matrix	corresponding correlation coefficient
1. G_jG_j'	
2. $JG_jG_j'J$	T–index (Janson & Vegelius, 1978a)
3. $G_jG_j'-n^{-1}\mathbf{11}'$	J–index (Janson & Vegelius, 1978a)
4. $G_jD_j^{-1}G_j'$	
5. $JG_jD_j^{-1}G_j'J$	T^2 coefficient (Tschuprow, 1939)
6. $2G_jG_j'-\mathbf{11}'-I$	Gamma coefficient (Hubert, 1977)

Some of these quantifications will now be discussed in more detail. That is, first the simplest quantification matrix, G_jG_j', will be shown to be a similarity matrix. Next, the fourth and fifth quantification matrices will be discussed, because these quantification matrices have often been adopted in practice, and will also be adopted in the second part of this study. The second, third, and sixth quantification matrices are not discussed, but can be interpreted in analogous ways.

3.2.1. The quantification matrix G_jG_j'

The elements of the quantification matrix G_jG_j' are given by

$$s_{ii'j} = \begin{cases} 1 \text{ if objects } i \text{ and } i' \text{ belong to the same category} \\ \\ 0 \text{ if objects } i \text{ and } i' \text{ belong to different categories,} \end{cases} \quad (1)$$

Clearly, $s_{ii'j}$ is a measure of similarity between objects i and i' in terms of variable j. That is, objects in the same category are seen as similar

($s_{ii'j} = 1$) and objects in different categories as dissimilar ($s_{ii'j} = 0$). This (binary) similarity measure is very simple, and does not take into account category frequencies or numbers of categories.

3.2.2. The quantification matrices $G_jD_j^{-1}G_j'$ and $JG_jD_j^{-1}G_j'J$

The quantification matrix $G_jD_j^{-1}G_j'$ is more complicated than the one discussed in the previous section. Let the category to which object i belongs be indicated by g, then

$$s_{ii'j} = \begin{cases} f_g^{-1} \text{ if objects } i \text{ and } i' \text{ belong to the same category} \\ \\ 0 \text{ if objects } i \text{ and } i' \text{ belong to different categories,} \end{cases} \quad (2)$$

where f_g is the g^{th} diagonal element of D_j, and thus the frequency of category g of variable j. Clearly, $s_{ii'j}$ can again be seen as a similarity measure, because its value is higher when objects fall in the same category (and hence are more similar) than when they belong to different categories. In contrast to the previous similarity measure, (1), the similarity between objects that fall in the same category now does depend on the number of objects that fall in this category. The more objects belong to this category, the less similar two objects that fall in this category are considered to be. Hence this measure in a way corrects for chance, because the higher the frequency of a category, the higher the probability that two objects would fall in it if the categories were statistically independent.

Although the similarity measure (2) seems to be attractive, it leads to a quantification matrix which has certain disadvantages. That is, the correlation between two qualitative variables, $\text{tr}\, S_j'S_l/(\text{tr}\, S_j^2)^{1/2}(\text{tr}\, S_l^2)^{1/2}$, with S_j and S_l as defined by (2) is always greater than zero, even when the variables are statistically independent. This follows from the fact that S_j and S_l are positive semi-definite (p.s.d.), hence $\text{tr}\, S_j'S_l \geq 0$ with equality if and only if the column-spaces of S_j and S_l are orthogonal. However, this equality can never be attained, because matrices $G_jD_j^{-1}G_j'$ and $G_lD_l^{-1}G_l'$ both contain the vector **1** in its column-space, and hence the column-spaces of S_j and S_l can never be orthogonal, as has been pointed out for instance by

Saporta (1975, p. IV–11). Saporta proposed to remedy this by centering the quantification matrix row– and column–wise. This leads to the quantification matrix $JG_jD_j^{-1}G_j'J$. It can be verified that the correlation between the thus defined quantification matrices is Tschuprow's T^2–coefficient (Tschuprow, 1939), which is a normalized version of the χ^2 measure. It is well–known that $\chi^2 = 0$ when two variables are statistically independent. Hence, the correlation between the thus defined quantification matrices for two statistically independent variables is 0. The elements of the quantification matrix can again be seen as similarities between the objects. They are now given by

$$s_{ii'j} = \begin{cases} f_g^{-1} - n^{-1} & \text{if objects } i \text{ and } i' \text{ belong to the same category} \\ -n^{-1} & \text{if objects } i \text{ and } i' \text{ belong to different categories.} \end{cases} \quad (3)$$

These similarities differ from those of (2) in that they are reduced by n^{-1}. This leads to negative similarities between objects that belong to different categories, and slightly reduced, but always positive, similarities between objects that fall into the same categories.

3.3. Quantification matrices for quantitative variables

For quantitative variables several quantification matrices can be used. Saporta (1976) and Janson and Vegelius (1982) both use the quantification matrix

$$S_j = n^{-1}\mathbf{z}_j\mathbf{z}_j' \quad (4)$$

where $\mathbf{z}_j$ is the vector of standardized scores on variable j. This quantification matrix is closely related to the scores that are ordinarily used in the analysis of quantitative variables. It can be interpreted as a similarity measure by noting that

$$s_{ii'j} = \begin{cases} n^{-1}|z_{ij}|\,|z_{i'j}| & \text{if the scores of } i \text{ and } i' \text{ have the same sign} \\ -n^{-1}|z_{ij}|\,|z_{i'j}| & \text{if the scores of } i \text{ and } i' \text{ have different signs.} \end{cases} \quad (5)$$

That is, two objects that have the same sign are seen as similar to a certain extent, while two objects with different signs are seen as dissimilar to a certain extent. The degree of (dis)similarity depends on the absolute values of the scores. Clearly, this measure of similarity only partly takes into account that objects with almost equal scores have higher similarity than objects with very different scores. The following similarity measure emphasizes this aspect of the similarity between two objects.

Gower (1971) proposed the following measure of similarity between objects with respect to a quantitative variable:

$$s_{ii'j} = 1 - |h_{ij} - h_{i'j}|/\rho_j \tag{6}$$

where h_{ij} is the score of object i on quantitative variable j, and ρ_j is the range of this variable. Clearly, this measure expresses the similarity between objects i and i', that is, the smaller the difference between the scores of the objects the higher the similarity. It is of interest to note that the matrix S_j with elements given by (6) is p.s.d., as Gower (1971) has shown.

Many other quantification matrices might be chosen for quantitative variables, for instance, by taking the outer product–moment of any of the vector quantifications for quantitative variables that are mentioned by Zegers (1986, pp. 34–41), among which the outer product–moment of the vector of raw scores and that of deviation scores.

3.4. Quantification matrices for ordinal variables

Although this study focuses on the analysis of qualitative variables and of mixtures of qualitative and quantitative variables, some attention needs to be paid to the analysis of ordinal variables. Defining a quantification matrix for ordinal variables allows one to analyze ordinal variables together with qualitative and quantitative variables.

As has been mentioned earlier, Daniels (1944) proposed to use certain particular square matrices as a kind of quantification matrices for ordinal variables. He has shown that Spearman's and Kendall's rank correlation coefficients can be formulated as normalized inner products between such quantification matrices. However, the "quantification matrices" he proposed are skew–symmetric matrices, while the methods described in chapter 2 apply to a set of symmetric matrices only. Therefore, these "quantification matrices" are not discussed here.

A very simple approach to handle ordinal variables is to treat the scores on such variables in the same way as quantitative variables. That is, one can use the quantification matrices that have been mentioned for the quantitative variables above with the ordinal scores considered as numerical scores (see Zegers, 1986, pp. 41–43). On the other hand, it is conceivable that other quantification matrices can be found that better describe the similarities between objects based on an ordinal variable. For instance, in the case of complete rank orders, that is, without ties, one may consider matrices with a simplex structure, that is, matrices with equal diagonal elements, and off–diagonal elements the size of which decreases as their distance from the diagonal increases. A particularly simple special case of such a simplex matrix is a tridiagonal matrix with unit elements on the diagonal and the two "by–diagonals", and zero elements elsewhere. This matrix could be generalized for the case of ties such that the similarity between objects in the same or neighbouring categories is set to 1, and the similarity between objects in more distant categories is set to 0. Obviously, these are only some examples of choices for quantification matrices for ordinal variables.

3.5. Normalization and weighting of quantification matrices

In methods that analyze a set of quantification matrices simultaneously, the quantification matrices may affect the solution differently. That is, some quantification matrices may affect the solution more than others, because they are "measured on a different scale". In order to prevent this one may weight the variables such that each affects the solution to the same extent. This situation parallels that of ordinary PCA, where variables can have different variances. For that reason, they are often normalized to constant sums of squares before a PCA is performed.

In order to normalize quantification matrices one needs to have an expression for how much a variable affects the solution. This can be obtained from the total sum of squares of the elements of the quantification matrix. Hence, in order to normalize a variable one may weight the quantification matrix by the inverse of the square root of the sum of squares of its elements.

Instead of weighting the variables such that they affect the solution equally one might want to do the opposite. That is, one might want to find a solution that maximally accounts for one variable, and analyzes the other variables only in the second place. This might be achieved by giving a large weight to that one variable and small weights to the others. This procedure has been proposed by Nishisato (1984) as "forced classification", in the MCA context. Yet another weighting procedure has been proposed by Cazes et al. (1976). They proposed to analyze a set of qualitative variables by means of MCA after weighting the quantification matrices $JG_jD_j^{-1}G_j'J$ for the variables by means of the elements of the first eigenvector of the (correlation) matrix of Tschuprow's T^2 coefficients between the variables.

3.6. Conclusion

Above, a number of quantification matrices for qualitative and quantitative variables have been described, and it has been indicated how these can be modified by normalizing or weighting them. These are only some of the quantification matrices that can possibly be used. In the present study no attempt is made to find the best choice for quantification matrices. However, some considerations that can be used for making this choice are discussed in section 4.4. For the purpose of the present chapter it suffices to mention that any quantification matrix may be used to represent a variable, as long as it is a symmetric matrix of order $n \times n$, that gives, in some way, similarities between the objects with respect to the variable concerned. The choice of the quantification matrix should be suitable for the data and research question at hand. This might imply that one has to invent similarity matrices oneself, or adopt ones that have been developed and presented in the abundant literature on similarity measures. On the other hand, one might adopt the most frequently made choices for quantification matrices, as made for instance by Saporta (1976), or adhere to the (technical) advices given by Janson and Vegelius (1982) or Zegers and Ten Berge (1986) for choosing quantification matrices that result in correlation coefficients with certain technical advantages.

4. A REVIEW OF THREE-WAY METHODS FOR THE ANALYSIS OF QUALITATIVE AND QUANTITATIVE TWO-WAY DATA

In the present chapter methods will be reviewed that result from applying the three–way methods that have been discussed in chapter 2 to the quantification matrices that have been discussed in chapter 3. It will be indicated which methods have been discussed elsewhere, and which methods appear to be new.

4.1. Three–way methods applied to quantification matrices for qualitative variables

In the present section a review is given of methods that result from applying the three–way methods discussed in chapter 2 to quantification matrices for qualitative variables (discussed in section 3.2). In Table 4.1 the three–way methods are crossed with the quantification matrices. The quantification matrices have been given without the j–indices for convenience. Division by "RSSQ", that is, by the square root of the sum of squares of the elements of the quantification matrix, indicates the normalized version of a quantification matrix. The cells that pertain to methods that have been discussed in the literature are filled with the names of those methods or their references.

The cross–classification of three–way methods and quantification matrices in Table 4.1 contains many empty cells. This is a consequence of the fact that the application of three–way methods to quantification matrices has not been studied in much detail yet, and as far as it has been studied, mostly the quantification matrix $JG_jD_j^{-1}G_j'J$ has been used. The methods in the first column, STATIS–1, come down to applying PCA to the quantification matrices belonging to the rows concerned. In many of their papers, Janson and Vegelius studied some aspects of these methods. In addition, they made a comparison of PCA of J–indices and Tschuprow's T^2–coefficients by means of an example data set (Janson & Vegelius, 1978b).

As has been described in chapter 2, PCA of quantification matrices for qualitative variables has also been developed in France, following Escoufier

Table 4.1. A cross-classification of methods for qualitative variables.

method quant. matrix	STATIS-1	TUCKALS-3	INDSCAL	INDORT	SUMPCA
GG'		Marchetti (1988)			
$\frac{JGG'J}{\text{RSSQ}}$	PCA of T-indices	Marchetti (1988)			
$\frac{GG'-n^{-1}11'}{\text{RSSQ}}$	PCA of J-indices				
$GD^{-1}G'$		Marchetti (1988)		INDOQUAL	MCA
$JGD^{-1}G'J$	PCA of ϕ^2-coef.	Marchetti (1988)		INDOQUAL	MCA
$\frac{JGD^{-1}G'J}{\text{RSSQ}}$	PCA of T^2-coef.	Marchetti (1988)		Kiers (1989c)	
$\frac{2GG'-11'-I}{\text{RSSQ}}$	PCA of H.'s Gamma				

(1970, 1973). In fact, the methods denoted here as PCA of T-indices and PCA of Tschuprow's T^2-coefficients have been proposed independently by Saporta (1975). The PCA of ϕ^2-coefficients (product-moment correlations for 2×2 contingency tables) has been proposed by Escoufier (1980).

Applying TUCKALS-3 to quantification matrices has been considered by Marchetti (1988). He mentions explicitly the quantification matrices G_jG_j' and $G_jD_j^{-1}G_j'$. In addition, he suggests that centering these matrices row- and column-wise, as well as scaling them, might be useful. That is, indirectly, he also suggests using $JG_jG_j'J$ and $JG_jD_j^{-1}G_j'J$, and normalized versions of these.

The third column of Table 4.1 is left empty, because applications of INDSCAL to a set of quantification matrices appear not to have been considered in any detail yet, although D'Ambra and Marchetti (1986) hint at it, and Kiers (1989c) also mentions it as a potentially useful method.

The application of INDORT to a set of quantification matrices has been described by Kiers (1989c), for the normalized versions of the quantification matrices $JG_jD_j^{-1}G_j'J$. He mentions that other quantification matrices might be chosen as well, but does not treat those in any detail. However, the application of INDORT to the (non-normalized) quantification matrices $JG_jD_j^{-1}G_j'J$ will be discussed in chapter 5, called "INDOQUAL" there. It will be noted that the INDOQUAL solution is closely related to that of INDORT applied to the quantification matrices $G_jD_j^{-1}G_j'$.

The last column contains only MCA. It is well-known that MCA can be seen as applying SUMPCA to a set of quantification matrices $JG_jD_j^{-1}G_j'J$ or $G_jD_j^{-1}G_j'$. The application of SUMPCA to quantification matrices for qualitative variables does not seem to have been considered in the literature. However, in the case of binary variables, SUMPCA has been applied to quantification matrices. Gower (1966) has discussed the application of SUMPCA to quantification matrices G_jG_j' for binary variables, albeit it not explicitly in this form. The methods for correspondence analysis of binary data discussed by Fichet (1986) and Fichet and Gbegan (1986) can be seen as particular variants of Gower's method. They have shown also that these methods are variants of ordinary MCA.

Apart from looking at the columns of Table 4.1 it is interesting to look at the rows of Table 4.1 as well. That is, a row of Table 4.1 in fact describes a hierarchy of methods for the analysis of qualitative variables, as follows from the fact that the three-way methods themselves are related hierarchically (see section 2.5). One of these hierarchies is the one for the quantification matrix $JG_jD_j^{-1}G_j'J$. That is, it can be concluded that MCA is a constrained variant of INDOQUAL (chapter 5), which is in turn a constrained variant of TUCKALS-3 applied to these quantification matrices (Marchetti, 1988). Finally, the latter method is a constrained variant of PCA of the quantification matrices $JG_jD_j^{-1}G_j'J$, that is, Escoufier's method of PCA of ϕ^2-coefficients.

Another interesting hierarchy is the one for the normalized version of the quantification matrix $JG_jD_j^{-1}G_j'J$. According to this hierarchy the method proposed by Kiers (1989c) is a constrained variant of one of the methods worked out by Marchetti (1988), which in turn is a constrained variant of PCA of Tschuprow's T^2-coefficients.

4.2. Three-way methods applied to quantification matrices for mixtures of qualitative and quantitative variables

In case one has a mixture of qualitative and quantitative variables one can again consider the application of the three-way methods from chapter 2 to the quantification matrices from chapter 3. The situation is a little more complicated than in the previous section, because now one has to choose quantification matrices both for the qualitative and the quantitative variables, which yields a large number of possible combinations. In fact one might make a three-way-table crossing quantification matrices for qualitative variables with quantification matrices for quantitative variables and with three-way methods. However, it seems that the quantification matrix $n^{-1}\mathbf{z}_j\mathbf{z}_j'$ is the most prevalent quantification matrix for quantitative variables. For this reason, only the slice of this three-way cross-classification that pertains to the quantification matrix $n^{-1}\mathbf{z}_j\mathbf{z}_j'$ for the quantitative variables, is given. The resulting two-way cross-table is given in Table 4.2. This second cross-classification can easily be related to the cross-classification given in Table 4.1, because the same three-way methods are crossed with the same quantification matrices. One might incorporate Table 4.1 as one slice of the three-way cross-table of quantification matrices for qualitative variables by quantification matrices for quantitative variables by three-way methods. Therefore, in the sequel reference is only made to one cross-classification which comprises methods for mixtures of qualitative and quantitative variables of which methods for merely qualitative variables are special cases.

The first column is again filled with a number of methods that have more or less explicitly been proposed by Janson and Vegelius (e.g., 1978a, 1982). That is, Janson and Vegelius discussed the correlation coefficients that are involved in these analyses, both for correlation between two qualitative variables, and for correlation between a qualitative and a quantitative variable. The methods are denoted here by the names Janson and Vegelius gave to the corresponding indices for correlation between a qualitative and a quantitative variable. An example of what is denoted here as "PCA of CP-indices" has been given by Janson and Vegelius (1982). Saporta (1976) also mentions the possibility of performing a PCA of mixed variables by means of PCA of what are called ZP-coefficients here.

Table 4.2. A cross-classification of methods for qualitative and quantitative variables.

method quant. matrix	STATIS-1	TUCKALS-3	INDSCAL	INDORT	SUMPCA
GG'					
$\frac{JGG'J}{RSSQ}$	PCA of SP-indices				
$\frac{GG'-n^{-1}11'}{RSSQ}$	PCA of CP-indices				
$GD^{-1}G'$				INDOMIX	PCAMIX
$JGD^{-1}G'J$				INDOMIX	PCAMIX
$\frac{JGD^{-1}G'J}{RSSQ}$	PCA of ZP-coef.			Kiers(1988)	
$\frac{2GG'-11'-I}{RSSQ}$					

The last column contains a method called PCAMIX in the present study. This method is a straight-forward generalization of PCA and MCA, such that it can handle mixtures of qualitative and quantitative variables. It has been proposed by many different authors independently, with slight variations, under names like Partially Optimal Scaling (Nishisato, 1980, p.103-107), or "Simultaneous treatment of qualitative and quantitative variables in factor analysis" (Escofier, 1979). This method is also contained as an option in PRINCALS (De Leeuw & Van Rijckevorsel, 1980). These methods have not been presented as applications of SUMPCA to quantification matrices, but it is readily verified that they can be written as such. For more details on these methods the reader is referred to chapter 7.

The other cells in the last column of Table 4.2 are left open. However, the application of the Gower (1966) method to similarities based on mixtures of qualitative and quantitative variables is likely to have been considered for many different similarity measures. Gower (1971) proposes his general association coefficient explicitly for the purpose of analyzing mixtures of qualitative and quantitative data. This similarity measure has been used in a discriminant analysis context by Cuadras (1989). As far as quantitative variables are concerned, both Gower and Cuadras use the quantification matrix defined by (6) in chapter 3. Therefore, these methods do not fit into the cross-classification of Table 4.2, but occur in a different slice of the three-way cross-classification that might be made.

Kiers (1988) has examined the application of INDORT to the normalized version of the quantification matrix $JG_jD_j^{-1}G_j'J$ for qualitative variables, but, in a practical example, he uses the non-normalized version. The latter method will be studied in detail in chapter 7, and is called "INDOMIX" there. Kiers (1988) mentions that other choices of quantification matrices might be made, and also that other methods might be used, like TUCKALS-3 and INDSCAL, but does not work these out.

As in Table 4.1, each row of Table 4.2 describes a hierarchy of three-way methods. That is, PCAMIX can be seen as a constrained variant of INDOMIX, which in turn is a constrained variant of PCA of the corresponding quantification matrices (that is $JG_jD_j^{-1}G_j'J$ for a qualitative variable and $n^{-1}\mathbf{z}_j\mathbf{z}_j'$ for a quantitative variable). The latter method comes down to PCA of a "correlation"-matrix of ϕ^2-coefficients for pairs of qualitative variables, squared product-moment correlations for pairs of quantitative variables, and η^2-coefficients for pairs of one qualitative and one quantitative variable.

4.3. Limitations of the given review

Above, a review has been given of methods for the analysis of mixtures of qualitative and quantitative variables, or sets of merely qualitative variables. It is by no means claimed that this review is exhaustive. The methods discussed here describe a class of methods that can be seen as applications of three-way methods to quantification matrices. One limitation of the review is that not all possible quantification matrices have been mentioned. Another limitation is that other three-way methods exist, that might be used for the analysis of a set of quantification matrices. The fact that only some three-way methods have been mentioned in Tables 4.1 and 4.2 is merely a matter of choice among the most familiar methods. This choice was partly based on the fact that the methods mentioned here could easily be seen to form a hierarchy. Other hierarchies are available as well, like the hierarchy mentioned by Carroll and Wish (1974).

Apart from the fact that only some three-way methods have been mentioned, a more important limitation of the review given here is that many methods for the analysis of qualitative variables cannot be considered as three-way methods applied to quantification matrices. The cross-classification does not comprise, for instance, the methods proposed by Domenges and Volle (1979), Escofier (1984), Lauro and D'Ambra (1984), Ter Braak (1986), Yanai (1986), Van der Heijden (1987), and Sabatier (1987), which treat the variables asymmetrically. The applications of three-way methods to quantification matrices can treat variables asymmetrically as well, by giving some variables a larger weight than others, as discussed in section 3.5. In the case of MCA, this procedure is equivalent to Nishisato's forced classification method (Nishisato, 1984). It is not clear, however, whether or not such methods yield results comparable to those of the methods discussed above.

Apart from the fact that the above review does not comprise the methods that treat variables asymmetrically, it does not contain all "symmetric" methods that have been developed for qualitative variables either. It does not, for instance, comprise the methods for multivariate analysis of qualitative (or mixed) variables developed by Young, Takane and de Leeuw (1978), Di Ciaccio (1986), Meulman (1986), Van Rijckevorsel (1987), Greenacre (1988) and Van der Burg (1988). Apart from these, methods for the analysis of qualitative variables in the linear structure analysis approach (for instance, Muthén, 1984) do not fit into this review either. The review can be said to be more or less complete, however, in that it seems to cover all methods that are known to form a compromise between the PCA-of-variables approach as proposed in the work of Janson and Vegelius, for instance, and the PCA-of-categories-and-objects approach of PCAMIX. If one wishes to perform a PCA of qualitative or mixed variables, then one has to decide whether one wants to perform PCA of the variables or of the categories and objects, because both of them at the same time is impossible. If one does not want to settle for either of them, one can use one of the compromise methods provided here, in order to achieve both objectives partly. Then it remains to choose the quantification matrix to be used, a choice which has not received much attention in the literature. In most cases one uses PCAMIX or related techniques for a PCA of mixtures of qualitative and quantitative variables, and MCA for a PCA of sets of qualitative variables. Thereby, one implicitly chooses one's quantification matrix. Although it is well–known that the quantification matrix used in MCA provides MCA with nice properties, it is by no means certain that using this quantification matrix yields the method adapted best to one's data.

4.4. How to choose one's method in practice

Above, a large number of methods for the analysis of qualitative and quantitative variables have been discussed. In addition, the idea of applying three–way methods to quantification matrices offers a practically unlimited number of new methods. There seems to be no ground for preferring one of these methods over all others. Empirical research might yield some comparative information on these methods, although it is not likely that conditions could be determined under which one of the methods is superior to all others. In the absence of empirical evidence the choice of the method is to be made by the practitioner, that is, on an ad hoc basis. In order to make such a choice a number of guidelines can be given that serve to clarify some implications of certain choices. These will be discussed now.

First of all the practitioner has to choose the quantification matrix that is to be used. In section 3.6 some remarks have been made concerning this choice. Having chosen one's quantification matrices, the next question to be answered is whether or not one wants to weight the variables in some way, including normalizing the variables. This question is difficult to answer, especially when one uses mixtures of qualitative and quantitative variables. A useful strategy might be based on the fact that the sum of squares of the elements of a quantification matrix indicates the (main) effect a variable has on the solution. That is, apart from the effect variables have on the solution because they are related more or less strongly (which might be called an interaction effect), there is an effect caused entirely by the size of the elements of the quantification matrix, and this is called the "main effect" here. One strategy one can adopt is to normalize the variables, thus ensuring that the main effects of the variables are equal, as is standard practice in

PCA. On the other hand, one might consider a qualitative variable to be more informative as it contains more categories. Then it seems desirable that the more categories a variable has, the more it affects the solution (in the sense of the main effect described above). Hence, in that case it would be better to use the non–normalized quantification matrices, the sums of squares of which are often directly related to the numbers of categories.

Although the above procedure may work for the analysis of merely qualitative variables, for the analysis of a mixture of qualitative and quantitative variables one cannot simply use this strategy, because one cannot compare a qualitative and a quantitative variable in terms of the numbers of categories they have. That is, when one decides to use non–normalized quantification matrices for the qualitative variables, one still has to decide what weights to attach to the quantitative variables. In one way, quantitative variables seem to be far more informative than qualitative variables, simply because of the fact that they make a finer distinction between the objects. If this interpretation of the informativeness of the variables seems appropriate for the data at hand, one may attach a very high weight to the quantitative variables, for instance, such that it has the same main effect as the qualitative variable with the highest main effect. On the other hand, one may consider a qualitative variable as a variable that has come about by combining several "latent" dimensions. For instance, one might consider a variable like "political preference" to be a combination of dimensions such as "conservative versus liberal" and "denominational versus non–denominational". In that case, it would be better to view a qualitative variable as a variable with more information than a quantitative variable, which therefore is to have more effect on the solution. A simple choice in that case would be to normalize quantification matrices for quantitative variables to unit sums of squares. This corresponds to the effect of a binary variable (that is, a qualitative variable with two categories) when it is quantified by $JG_jD_j^{-1}G_j'J$. This is one of the decision problems which might be alleviated when comparative empirical results on these two strategies are available.

Apart from the choice of a quantification matrix, one has to choose the three–way method to apply to the quantification matrices. If there is no a priori reason for using one and only one of the three–way methods from chapter

2, then a useful strategy might be to use several of them, and decide afterwards which method yields the best interpretable description of the data. In making such decisions, at least two considerations have to be borne in mind. First, it is important to note that STATIS–1 can give the best representation of the variables only at the cost of giving no representation at all of the objects. If one is not interested in the representation of the objects at all, then this might be the best method to use. However, as soon as one wants to have descriptions of both objects and variables, then STATIS–1 is no longer useful. The second consideration is that there is a trade–off in the adequacy of the description of the variables and the complexity of the model that is used. A strategy might be to choose between the different methods by choosing that method that is the lowest in the hierarchy, and hence the simplest, that still gives a reasonable representation of the variables, as has been suggested in section 2.6.

The guidelines given above only partly help someone who is to analyze qualitative or mixed variables by means of PCA. Obviously, many problems of choice between methods in the above review are yet to be investigated. It has been the aim of the present section, however, to indicate what questions can be posed, and how one might answer these.