PART III

ANALYSES OF EMPIRICAL DATA

10. EXPERIENCES WITH INDOQUAL AND INDOMIX

The present chapter contains several example analyses. Each of the examples have been chosen for a special purpose. The first example (section 10.2) deals with the classification of whales. As has been mentioned in section 8.7, both Multiple Correspondence Analysis (MCA) and INDOQUAL can be used as techniques for clustering objects on the basis of qualitative variables, while INDOQUAL yields a slightly better solution. This first example analysis has been treated in much detail, in order to demonstrate how one can interpret the results from an INDOQUAL analysis. In addition, a stability analysis has been performed on the MCA and INDOQUAL solutions for this data set.

In the second example (section 10.3) the results of an enquiry reported by Vegelius and Bäckström (1981) are analyzed with the purpose of finding the most important components underlying this enquiry. In this example, the emphasis is on the variables. It is demonstrated, however, that focusing on the variables *only*, as PCA of quantification matrices does, may be too limited. The results from an INDOQUAL analysis, which considers the objects and categories additionally, are discussed.

In section 10.4 an example is given of an INDOMIX analysis. The data from a survey on abortion and related issues (described in Gifi, 1981) are analyzed here. The analysis uses a special weighting option in order to control for background variables. The variables are a mixture of variables that can be considered as nominal and ordinal variables. The stability of the solution is assessed by means of cross–validation.

The fourth example (section 10.5) describes the analysis of a set of binary variables, reported by Van Zomeren and Van den Burg (1985). The purpose of this analysis is, again, to find certain subscales or clusters of variables. Some of the special possibilities of the analysis of binary variables have been used.

In order to demonstrate to what extent standardizing the variables can affect the INDOQUAL solution, the fifth example data set (section 10.6) has been analyzed both after standardizing the variables, and without standardizing the variables, the latter being the ordinary INDOQUAL procedure. The data consists of two binary and five nominal variables with

more than three categories. It appears that standardizing the variables results in a domination of the solution by the binary variables.

The sixth example (section 10.7) has been taken up in order to compare the results of an INDOQUAL solution with those of a TUCKALS–3 analysis of quantification matrices, as proposed by Marchetti (1988). The data have been analyzed by INDOQUAL with the same options as chosen by Marchetti (1988), and it is shown that the results are very similar.

Finally, the seventh data set has been analyzed to give an example of what may go wrong in blindly applying INDOQUAL to any data set of nominal variables (section 10.8). It is shown that, because the variables in the data set at hand (Sugiyama, 1975) do not form any clusters of closely related variables at all, as can be verified by means of studying the generalized correlations among them, each component of the INDOQUAL solution is determined by one and only one variable. Obviously, such a solution is not at all interesting from the point of view of data reduction.

As has been mentioned above, the stability of the solutions will be considered in two cases. In the next section, the methods will be described by means of which these aspects of the stability are assessed.

10.1. Assessing the stability of INDOQUAL and INDOMIX solutions

All methods for the multivariate analysis of qualitative and quantitative variables described here are usually applied to samples from a population. In principle, one attempts to find a description of the data that is valid for the larger population from which the sample is drawn. Observations on the complete population are rarely available. Hence it is important to know whether or not the sample is representative for the population. One way of determining this is to draw another sample. However, in practice this often is not possible. In order yet to determine how sensitive the results of an analysis are to the properties of the particular sample one may analyze the stability of the solution over deletion of certain observations, or, alternatively, one may cross–validate the results from one half of the sample by means of the other half of the sample. These two methods will be discussed briefly in sections 10.1.1, and 10.1.2, respectively.

10.1.1. Stability over deletion of certain observations (jackknifing)

The first procedure for studying how much the results of an analysis depend on the particular sample is based on alternatingly analyzing the data with one object left out. This procedure is called the "jackknife" technique (e.g., Miller, 1974), or simply "jackknifing". In INDOMIX (and hence in INDOQUAL) jackknifing will be done as follows. When the number of objects is small, the data will be analyzed n times, with each object deleted once. Each sample with one object deleted is called a jackknife. This procedure is called "complete jackknifing", because each object is eliminated once. When the number of objects is large, "random jackknives" will be used. That is, a fixed number of jackknives will be determined by leaving out one randomly determined object each time. Next, the results of the analyses of the different jackknives are compared to the results of the analysis of the original sample and to each other, in order to assess the stability of the original solution.

In both complete and random jackknifing one obtains a large number of results of INDOMIX analyses on the different jackknives. The basic elements of the jackknife solutions are the object scores and the loadings for the variables. The object scores, however, cannot all be compared over the different jackknife solutions, because in each jackknife solution partly different sets of objects are used. On the other hand, the measures called category centroids, which are based directly on the objects, can be compared across all different jackknife solutions. All jackknife solutions contain loadings and category centroids for the same variables and categories, respectively.

The purpose of the jackknifing procedure used here is to determine to what degree the separate jackknife solutions differ from each other and from the original solution. Hence one has to compare a large number of loadings and category centroids over all jackknives. As has been explained by De Leeuw and Meulman (1986), comparing different jackknife solutions in multidimensional scaling methods requires that the solutions be matched to each other, before one can compare solutions. Because the INDORT solution yields unique axes, such matchings are unnecessary for comparing results from different INDOMIX analyses.

A useful way of comparing the loadings and category centroids (both

denoted as "parameters" here) of different jackknife solutions with each other seems to be to compute the means and standard deviations of the jackknife parameters. The means over the jackknife studies can be compared with the observed scores in the original study in order to see if the jackknife results differ systematically from the original study. The standard deviations give information on the stability of the parameters over the different jackknives. The stability results provided here are by no means given for hypothesis testing. They merely serve to indicate how stable the solution is as a whole, and, specifically, how stable each of the individual parameters is.

The computation of an INDOMIX solution requires an iterative process, which is rather time consuming. In the jackknifing procedure sketched above, this procedure is to be repeated a number of times, that is, as many times as there are objects to be deleted. In addition, each INDOMIX analysis should be repeated a number of times in order to check whether or not the global optimum has been found. All this would require large computation times. However, the iterative process can be accelerated by using good start configurations. If deleting an object does not cause dramatic changes in the solution, it can be assumed that the solution of one jackknife will provide a good starting configuration for the iterative process for computing another jackknife solution. Therefore, the different jackknife solutions have been computed in this way, that is, using the solution from the previously computed jackknife in order to find the new jackknife solution.

10.1.2. Cross-validation via a split-half procedure

Apart from jackknifing, another procedure is considered for determining the dependence of a solution on particular characteristics of the sampled data. This procedure is "cross-validation" via a split-half procedure. That is, first, one randomly splits the data into two halves, and the first half of these is analyzed by means of INDOMIX. This analysis yields weights, to be explained later, for the categories of qualitative variables and for the quantitative variables. These weights are used to compute object scores for the data in the second half. These "pseudo" object scores are compared to the "original" object scores (resulting from INDOMIX on the second half) by inspecting the correlation between them. Finally, one can compute how well these object scores represent the data in the second half, by simply computing

the loadings of the object scores on the variables. Obviously, one may, in addition use the inverse procedure, that is, cross–validating the INDOMIX results of the analysis of the second half by applying the resulting weights to the variables in the first half.

The weights to be used for computing object scores from the variables are based on the following. It has been mentioned in section 9.3 that the INDOMIX object scores can be written as $X = UB$, for some matrix B of weights. These weights can be applied to any data set for which a U is given with columns referring to the same quantitative variables and/or categories of qualitative variables. In fact, these weights resemble the "component weights" in ordinary PCA, which are used to compute the component scores as linear combinations of the variables. In the present case these weights can be used to compute the object scores as linear combinations of the columns of U, that is, of the standardized quantitative variables and the columns of the transformed indicator matrices for the qualitative variables. It should be noted that the resulting object scores are not necessarily uncorrelated. Therefore, it is interesting to compute the correlation between these components as well.

10.2. The cetacea data: MCA and INDOQUAL as clustering techniques

Vescia (1985b) has collected data on 36 cetacea (whales, porpoises and dolphins) on the basis of zoological descriptions. The cetacea have been "measured" on 15 variables, describing morphological, osteological, and behavioral aspects of the animals under study. These variables have been described in detail by Vescia (1985b), as well as by Meulman (1986, pp.28–33), albeit in a different order. There were a few missing observations. Meulman (1986) has considered a missing observation on a variable as "falling in a different category". This could be justified by the fact that most of the (few) missing observations occurred systematically within one or two families of cetacea. That is, "missing" may be considered as a characteristic in its own right. In the present study missing data have been handled in the same way. The data have been analyzed essentially as they have been given (in their coded form) by Meulman (1986), except that one accidentally omitted white whale has been recovered, and the four errors that had emerged in the original data set, as pointed out by Vescia (1985b, p.13), have been corrected. One of the "missing" data was in fact based on a coding error. The data set analyzed

here is given in Table 10.1, where the corrected data are printed in bold face. As can be verified by comparing the MCA results reported by Meulman (1986) and those given here, the errors hardly affected the solution.

Table 10.1. *The cetacea data.*

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	family
1	6	1	1	1	2	3	1	4	5	2	3	1	3	4	baleen whale
1	6	1	1	1	2	3	1	4	2	2	3	1	3	4	baleen whale
1	2	1	1	4	1	3	1	4	1	2	3	1	3	4	baleen whale
1	2	1	1	1	4	4	1	4	4	1	3	1	3	4	grey whale
1	5	1	1	4	1	5	3	4	1	1	3	1	5	4	finback whale
1	5	1	1	4	4	5	3	4	4	1	3	1	5	4	finback whale
1	5	1	1	4	4	5	3	4	4	1	3	1	**5**	4	finback whale
1	1	1	1	2	1	1	1	1	1	**2**	3	4	5	1	sperm whale
1	1	1	1	1	1	1	1	1	1	**2**	1	4	2	1	sperm whale
2	3	2	3	3	1	1	2	3	1	2	2	3	3	1	beaked whale
2	3	2	3	3	1	1	2	3	1	2	2	3	5	1	beaked whale
2	2	2	3	3	1	1	2	3	1	2	2	3	5	1	beaked whale
1	4	2	3	3	1	1	2	3	1	2	2	3	3	1	beaked whale
1	2	2	3	3	1	1	2	3	4	2	2	3	5	1	beaked whale
1	2	2	2	3	3	2	1	2	5	2	2	2	3	1	dolphin
1	3	2	3	3	3	2	1	2	1	2	2	2	2	2	dolphin
1	4	2	1	3	4	2	1	2	4	2	2	2	2	1	dolphin
1	4	2	1	3	4	1	1	2	1	2	2	2	5	1	dolphin
1	3	2	3	3	3	2	1	2	1	2	2	2	3	2	dolphin
1	2	2	2	1	3	2	1	2	1	2	2	2	2	2	dolphin
2	4	2	1	3	1	2	1	2	2	2	2	2	4	2	dolphin
1	4	2	1	3	2	2	1	2	**4**	2	2	2	3	3	dolphin
1	4	2	1	3	4	2	1	2	2	2	2	2	5	1	dolphin
1	3	2	3	3	3	2	1	2	3	2	2	2	4	2	dolphin
1	3	2	3	3	3	2	1	2	1	2	2	2	4	2	dolphin
1	3	2	3	3	3	2	1	2	4	2	2	2	5	2	dolphin
1	2	2	3	3	3	2	1	2	1	2	2	2	2	2	dolphin
1	3	2	3	3	3	2	1	2	1	2	2	2	5	2	dolphin
1	2	2	1	1	1	2	1	2	4	2	2	2	5	2	porpoise
1	2	2	1	2	1	2	1	2	1	2	2	2	5	2	porpoise
2	4	2	1	1	2	2	1	2	3	1	2	3	3	2	white whale
2	**4**	**2**	**1**	**1**	**2**	**1**	**1**	**2**	**4**	**1**	**2**	**3**	**3**	**1**	white whale
2	3	2	4	2	2	2	1	3	1	1	1	5	1	2	river dolphin
2	3	2	4	2	2	2	1	1	1	1	1	5	1	2	river dolphin
2	3	2	4	2	2	2	1	3	2	1	1	2	1	2	river dolphin
2	3	2	4	2	2	2	1	3	2	1	1	5	1	2	river dolphin

The cetacea can be classified into several families. For each of the cetacea, the family name is given in Table 10.1. These families can be

grouped hierarchically into several classes. According to the theory of Grasse (1955, see also Vescia, 1985a, p.16; Meulman, 1986, p. 29) the classification given in Figure 10.1 can be made.

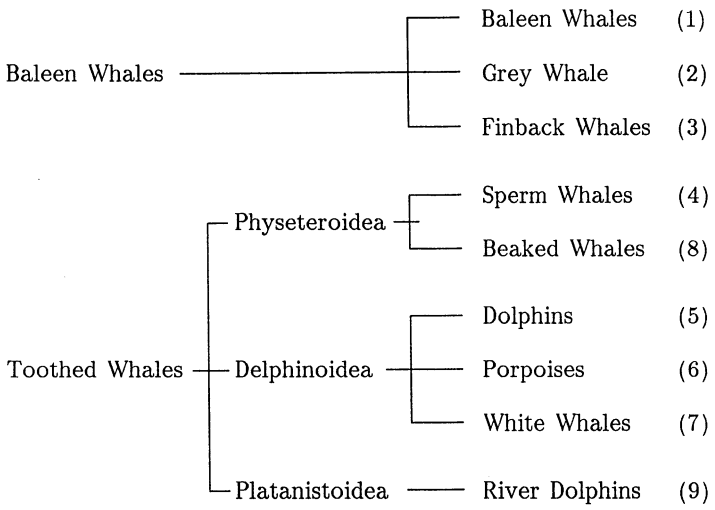

Figure 10.1. *The classification of cetacea according to Grasse.*

In the present study the data have been analyzed by means of both MCA and INDOQUAL. The former analysis has also been reported by Van der Burg (1985) and Meulman (1986, pp. 28–33). Both show that MCA (or Homogeneity analysis, as it is called there) is to a certain extent capable of distinguishing the original 9 families of cetacea, which is in accordance with Van der Burg's (1988) remark that MCA can be seen as a clustering technique. As has been described in chapter 8, INDOQUAL may be expected to yield clusters that are more compact and more clearly separated than the ones resulting from MCA. Therefore, the MCA solution will be compared to the INDOQUAL solution.

In the present analyses the two–dimensional solutions of MCA (as in the earlier analyses) and INDOQUAL are considered. This choice has been based partly on the fact that after the second component adding extra components did not increase much the amount of inertia accounted for by INDOQUAL. That is, the inertias accounted for by the one–, two–, three–, four–, and

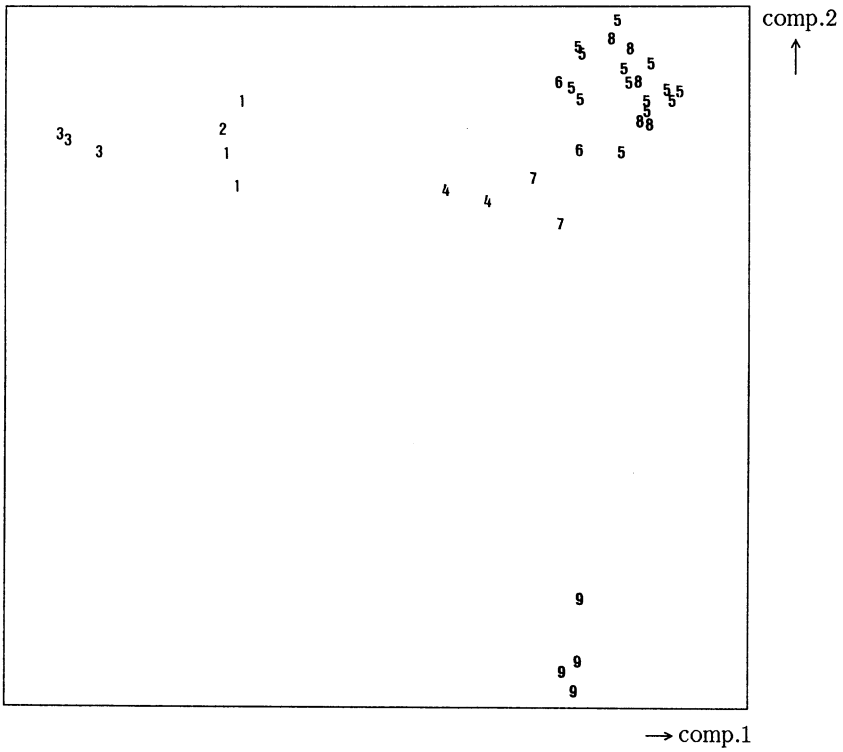

Figure 10.2. *The object coordinates resulting from MCA.*

five–dimensional INDOQUAL solutions are 6.9 (16%), 11.2 (26%), 14.1 (33%), 16.6 (38%), and 19.0 (44%), respectively. Clearly, the percentages are quite low, indicating that much of the information remains unaccounted for. One may be satisfied with a small amount of inertia accounted for, however, when one is interested in the most predominant relations between the variables. Because the two–dimensional INDOQUAL solution was well interpretable in terms of the original families of cetacea, and distinguished most of the families very well, it was decided to settle for this solution in the present study.

The two–dimensional MCA solution has been rotated to simple structure by means of a varimax rotation. The angle of rotation was 6.3°, which implies that, in this case, the principal axes solution of MCA had almost optimal simple structure.

The main purpose of the present analysis is the comparison of the object coordinates resulting from MCA and INDOQUAL. These have been plotted in Figures 10.2 and 10.3, respectively. The cetacea have been coded by their

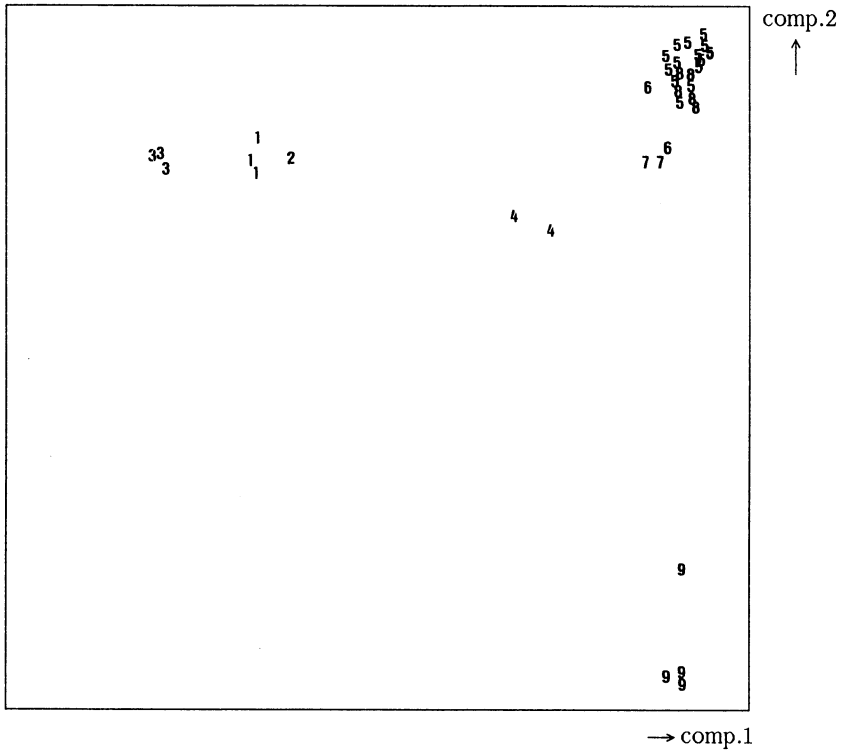

Figure 10.3. *The object coordinates resulting from INDOQUAL.*

family labels. In both plots, the families 1, 2, 3, 4, and 9, are well separated. The families 5, 6, 7, and 8 are not far apart. Moreover, the families 1, 2, and 3 are not far apart, and well separated from the other families, as would be expected from Grasse's classification. Similarly, the closeness of families 5, 6, and 7 is in agreement with Grasse's classification. Families 4 and 9 are both quite far from the other families, again in line with Grasse's classification. On the other hand, families 4 and 8 are separated much more than one would expect on the basis of Grasse's classification. The fact that family 8 is found within the cluster of families 5, 6, and 7 is unexpected as well. The results from MCA and INDOQUAL are globally equivalent. The differences between the MCA solution and the INDOQUAL solution are to be found in the details. It can be observed that in the INDOQUAL solution the families 1 and 2 are distinguished better than in the MCA solution. A similar result holds for families 5 and 6. The families 5 and 8 are completely intertwined in the MCA solution, whereas in the

INDOQUAL solution the members of family 8 are more to the bottom of the cluster with members of family 5, and "disturbed" only by a few of these members. Finally, the members of family 7 are closer to each other in the INDOQUAL solution than in the MCA solution. These slight differences all indicate that in the INDOQUAL solution the families of cetacea are distinguished better than in the MCA solution. Incidentally, it should be noted that the solution obtained by Meulman (1986, p.127) via a method that fits the distances between profiles directly yields results that are yet better interpretable in terms of the original families of cetacea. This was to be expected because her method has been designed specifically for the purpose of representing (distances between) objects. INDOQUAL on the other hand has been developed for the purpose of representing both objects and variables.

Table 10.2. *Loadings of the variables for MCA and INDOQUAL.*

	MCA + varimax		INDOQUAL	
	comp.1	comp.2	comp.1	comp.2
1. Neck	0.10	0.36	0.10	0.32
2. Form of the head	**0.80**	0.22	**0.79**	0.17
3. Size of the head	**0.82**	0.00	**0.86**	0.00
4. Beak	0.40	**0.95**	0.32	**0.96**
5. Dorsal fin	**0.77**	**0.69**	**0.76**	**0.71**
6. Flippers	0.27	0.49	0.19	0.46
7. Set of teeth	**0.94**	0.05	**0.97**	0.02
8. Longitudinal furrows	0.56	0.05	0.51	0.03
9. Blow hole	**0.93**	0.29	**0.97**	0.35
10. Color	0.13	0.17	0.09	0.13
11. Cervical vertebrae	0.17	0.45	0.13	0.42
12. Lachrymal & jugal bones	**0.87**	**0.81**	**0.91**	**0.84**
13. Head bones	**0.93**	**0.74**	**0.98**	**0.78**
14. Habitat	0.15	**0.92**	0.12	**0.93**
15. Feeding	**0.92**	0.13	**0.95**	0.08
sum of loadings	8.76	6.32	8.65	6.20
sums of squares of loadings	6.73	4.16	6.92	4.28

Next, the loadings resulting from MCA (followed by varimax rotation) are compared to those of INDOQUAL. These loadings are given in Table 10.2, with loadings $\geq$ .65 printed in bold face. The loadings from MCA and INDOQUAL are highly similar. Nevertheless, there is a tendency for loadings that are high in the MCA solution to correspond to even higher loadings in the INDOQUAL solution, and for loadings that are small in the MCA solution to correspond to even smaller loadings in the INDOQUAL solution. Table 10.2 also gives the inertia accounted for by the components in terms of the MCA–model (sums of loadings) and in terms of the INDOQUAL–model (sums of squares of the loadings).

From Table 10.2 it is clear which variables are important for the interpretation of each of the components. It is, however, not clear from these loadings how the variables are related to the components. In order to get more specific results it is useful to consider the category coordinates for all variables. These are computed as the means of the object scores of objects that belong to the category concerned. The category coordinates for the INDOQUAL solution are given in Table 10.3.

Using the results given in Tables 10.2 and 10.3, one can interpret the two components as follows. On the first component the variables 2, 3, 5, 7, 9, 12, 13 and 15 have high loadings (greater than 0.65). Therefore, this component will be interpreted in terms of the categories of these variables only. The first component can be interpreted by means of the following contrasts:

flat or convex heads	others	(variable 2)
very big heads	medium sized heads	(variable 3)
backward and falciform fin	other or no fin	(variable 5)
without teeth	with teeth	(variable 7)
double blow hole	single blow hole	(variable 9)
missing observation	other categories	(variable 12)
symmetrical headbones	asymmetrical headbones	(variable 13)
plankton eaters	others	(variable 15)

This first component is also the component that contrasts the super family of the Baleen Whales to that of the Toothed Whales, which is in agreement with the interpretation in terms of category coordinates given above.

On the second component the variables 4, 5, 12, 13, and 14 have high

loadings. In terms of the categories of these variables this component can be interpreted as follows:

narrow and long beak	other beak	(variable 4)
triangular fin	other or no fin	(variable 5)
linked lachrymal & jugal bones	independent bones	(variable 12)
missing observation	other categories	(variable 13)
river dwellers	others	(variable 14)

As is clear from Figure 10.3, this second component especially contrasts river dolphins, family 9, to the other cetacea, which corresponds well to the interpretation given above. This finishes the interpretation of the results of INDOQUAL on the cetacea data.

The stability of the INDOQUAL solution for the cetacea data has been examined by means of a jackknife study, as has been explained in section 10.1.1. That is, 34 analyses (both MCA with varimax rotation and INDOQUAL) have been performed on these data with each time one of the animals left out. Because leaving out the fourth or 22^{nd} cetacean would result in empty categories, no analyses have been performed with these animals left out. The 34 analyses resulted in sets of loadings and category coordinates that are comparable over the analyses. In order to assess the stability of all the parameters the standard deviations over the 34 jackknives have been computed. These are given, together with the means over the 34 jackknives, in Table 10.4 for the loadings, and in Table 10.5 for the category coordinates.

Comparing these results with those of the original analyses on the complete data set shows that the means of the loadings and category coordinates overthe jackknives are very close to the loadings and category coordinates in the original study. More importantly, the standard deviations over the jackknives are typically very small both in the MCA and INDOQUAL results. It can be seen that in particular the high and small loadings of INDOQUAL tend to be more stable than those in MCA, whereas the medium sized loadings tend to be more stable in MCA.

In order to indicate how one may translate the size of the standard deviation in a measure for stability, it is useful to remark that the standard deviation can be interpreted as a kind of mean deviation from the mean of the parameter value over the jackknife solutions. Instead, the true mean of

Table 10.4. *Means and standard deviations of the loadings over* 34 *jackknife studies.*

	MCA + varimax		INDOQUAL	
	comp.1	comp.2	comp.1	comp.2
var.1	0.10 (0.009)	0.36 (0.030)	0.10 (0.008)	0.32 (0.032)
var.2	0.80 (0.016)	0.22 (0.022)	0.79 (0.021)	0.17 (0.019)
var.3	0.82 (0.015)	0.00 (0.000)	0.86 (0.016)	0.00 (0.000)
var.4	0.40 (0.019)	0.95 (0.007)	0.32 (0.020)	0.96 (0.003)
var.5	0.77 (0.017)	0.68 (0.035)	0.76 (0.021)	0.71 (0.034)
var.6	0.28 (0.030)	0.48 (0.025)	0.20 (0.025)	0.46 (0.032)
var.7	0.94 (0.004)	0.05 (0.017)	0.97 (0.004)	0.02 (0.006)
var.8	0.56 (0.036)	0.05 (0.014)	0.50 (0.041)	0.03 (0.006)
var.9	0.93 (0.004)	0.29 (0.038)	0.97 (0.003)	0.36 (0.025)
var.10	0.13 (0.024)	0.18 (0.033)	0.09 (0.019)	0.14 (0.027)
var.11	0.17 (0.026)	0.44 (0.031)	0.13 (0.024)	0.42 (0.033)
var.12	0.87 (0.012)	0.81 (0.029)	0.91 (0.013)	0.85 (0.023)
var.13	0.94 (0.004)	0.74 (0.043)	0.98 (0.002)	0.77 (0.041)
var.14	0.15 (0.013)	0.92 (0.009)	0.12 (0.011)	0.93 (0.006)
var.15	0.91 (0.005)	0.13 (0.026)	0.95 (0.005)	0.08 (0.013)

Table 10.5. *Means and standard deviations of the category coordinates over* 34 *jackknife studies.*

	MCA + varimax		INDOQUAL	
var,cat	comp.1	comp.2	comp.1	comp.2
1,1	−0.19 (0.013)	0.37 (0.016)	−0.19 (0.013)	0.35 (0.019)
1,2	0.50 (0.019)	−0.97 (0.060)	0.51 (0.018)	−0.92 (0.061)
2,1	−0.27 (0.055)	−0.16 (0.120)	−0.30 (0.050)	−0.32 (0.054)
2,2	0.04 (0.047)	0.51 (0.021)	0.01 (0.050)	0.41 (0.022)
2,3	0.65 (0.021)	−0.61 (0.049)	0.58 (0.020)	−0.53 (0.045)
2,4	0.35 (0.022)	0.37 (0.033)	0.50 (0.018)	0.37 (0.022)
2,5	−2.42 (0.093)	0.05 (0.026)	−2.31 (0.087)	0.06 (0.012)
2,6	−1.57 (0.087)	−0.06 (0.051)	−1.78 (0.077)	0.04 (0.014)
3,1	−1.56 (0.056)	0.02 (0.023)	−1.60 (0.056)	−0.01 (0.011)
3,2	0.52 (0.017)	−0.01 (0.008)	0.53 (0.018)	0.00 (0.004)
4,1	−0.63 (0.027)	0.17 (0.022)	−0.57 (0.028)	0.17 (0.017)
4,2	0.55 (0.032)	0.73 (0.047)	0.50 (0.021)	0.65 (0.032)
4,3	0.70 (0.022)	0.52 (0.025)	0.61 (0.020)	0.55 (0.024)
4,4	0.47 (0.020)	−2.71 (0.140)	0.49 (0.019)	−2.73 (0.151)
5,1	−0.47 (0.066)	0.09 (0.034)	−0.47 (0.069)	0.12 (0.025)
5,2	0.31 (0.026)	−1.80 (0.097)	0.34 (0.028)	−1.84 (0.097)
5,3	0.59 (0.019)	0.53 (0.020)	0.58 (0.019)	0.54 (0.024)
5,4	−2.19 (0.090)	0.12 (0.015)	−2.17 (0.088)	0.09 (0.009)

Table 10.5. *Means and standard deviations of the category coordinates over* 34 *jackknife studies (continued).*

	MCA + varimax		INDOQUAL	
var,cat	comp.1	comp.2	comp.1	comp.2
6,1	−0.00 (0.055)	0.29 (0.048)	−0.03 (0.047)	0.23 (0.026)
6,2	−0.06 (0.053)	−1.19 (0.050)	−0.03 (0.055)	−1.14 (0.056)
6,3	0.70 (0.024)	0.55 (0.029)	0.60 (0.021)	0.64 (0.028)
6,4	−0.95 (0.102)	0.37 (0.031)	−0.79 (0.099)	0.30 (0.023)
7,1	0.34 (0.038)	0.34 (0.070)	0.34 (0.032)	0.23 (0.034)
7,2	0.53 (0.019)	−0.18 (0.034)	0.54 (0.018)	−0.13 (0.018)
7,3	−1.55 (0.083)	0.07 (0.043)	−1.77 (0.076)	0.10 (0.012)
7,4	−1.62 (0.058)	0.15 (0.019)	−1.59 (0.065)	0.09 (0.013)
7,5	−2.42 (0.093)	0.05 (0.026)	−2.31 (0.087)	0.06 (0.012)
8,1	0.15 (0.023)	−0.10 (0.018)	0.14 (0.022)	−0.08 (0.010)
8,2	0.63 (0.030)	0.53 (0.074)	0.57 (0.022)	0.42 (0.031)
8,3	−2.42 (0.093)	0.05 (0.026)	−2.31 (0.087)	0.06 (0.012)
9,1	−0.04 (0.062)	−1.03 (0.256)	−0.06 (0.066)	−1.17 (0.198)
9,2	0.50 (0.020)	0.44 (0.024)	0.54 (0.018)	0.49 (0.023)
9,3	0.58 (0.022)	−0.67 (0.106)	0.55 (0.019)	−0.73 (0.088)
9,4	−1.93 (0.071)	0.07 (0.013)	−1.97 (0.072)	0.08 (0.008)
10,1	0.23 (0.035)	0.02 (0.046)	0.18 (0.031)	−0.00 (0.036)
10,2	0.06 (0.080)	−0.96 (0.148)	0.07 (0.084)	−0.84 (0.135)
10,3	0.58 (0.079)	0.12 (0.107)	0.54 (0.034)	0.36 (0.077)
10,4	−0.50 (0.069)	0.36 (0.022)	−0.37 (0.063)	0.31 (0.016)
10,5	−0.54 (0.294)	0.45 (0.126)	−0.63 (0.283)	0.37 (0.080)
11,1	−0.66 (0.058)	−1.07 (0.049)	−0.58 (0.059)	−1.05 (0.052)
11,2	0.25 (0.022)	0.41 (0.021)	0.22 (0.023)	0.40 (0.021)
12,1	0.34 (0.026)	−2.20 (0.118)	0.35 (0.029)	−2.25 (0.118)
12,2	0.53 (0.018)	0.46 (0.020)	0.54 (0.018)	0.48 (0.022)
12,3	−1.74 (0.064)	0.05 (0.014)	−1.78 (0.064)	0.03 (0.009)
13,1	−1.93 (0.071)	0.07 (0.013)	−1.97 (0.072)	0.08 (0.008)
13,2	0.54 (0.020)	0.34 (0.041)	0.55 (0.019)	0.38 (0.034)
13,3	0.51 (0.030)	0.34 (0.066)	0.52 (0.021)	0.32 (0.030)
13,4	−0.27 (0.055)	−0.16 (0.120)	−0.30 (0.050)	−0.32 (0.054)
13,5	0.46 (0.021)	−2.82 (0.155)	0.48 (0.019)	−2.87 (0.163)
14,1	0.47 (0.020)	−2.71 (0.140)	0.49 (0.019)	−2.73 (0.151)
14,2	0.45 (0.039)	0.46 (0.046)	0.40 (0.031)	0.42 (0.053)
14,3	−0.27 (0.058)	0.26 (0.027)	−0.30 (0.055)	0.27 (0.017)
14,4	0.76 (0.031)	0.37 (0.044)	0.63 (0.022)	0.60 (0.034)
14,5	−0.26 (0.051)	0.35 (0.027)	−0.20 (0.047)	0.31 (0.021)
15,1	0.35 (0.029)	0.43 (0.049)	0.39 (0.026)	0.32 (0.028)
15,2	0.57 (0.021)	−0.38 (0.043)	0.54 (0.019)	−0.30 (0.028)
15,3	0.31 (0.027)	0.44 (0.031)	0.54 (0.019)	0.37 (0.025)
15,4	−1.93 (0.071)	0.07 (0.013)	−1.97 (0.072)	0.08 (0.008)

absolute deviations from the mean parameter value might have been used. The standard deviation has been chosen here because it is more sensitive to extreme deviations than the mean absolute deviation. When the standard deviation is small, this does not only imply that the deviations are small on the average, but also that extreme deviations from the mean parameter value do not occur.

It can be concluded that both MCA and INDOQUAL yield two–dimensional solutions that can be interpreted very well. The solutions differ slightly in that the INDOQUAL solution tends to give slightly more compact and more clearly separated clusters of cetacea families. In addition, most parameters in both solutions are very stable.

10.3. An enquiry about religion: Components analysis of nominal variables

Vegelius and Bäckström (1981) have analyzed the results of an enquiry among 118 theological students. The enquiry contained 24 questions on social and religious background (demographic variables), religious activities, plans for the future, and attitudes towards miscellaneous issues. All variables are considered as nominal variables. The variables are mentioned in Table 10.6, for the categories of the variables the reader is referred to Vegelius and Bäckström. Vegelius and Bäckström (1981) performed a PCA on the matrix of Tschuprow's T^2 coefficients among the variables. This method has been explained in section 4.1. Their analysis yielded 8 eigenvalues greater than one, and for this reason they reported 8 components. The loadings for the variables have been rotated to simple structure by means of a varimax rotation. This solution had a clear simple structure, and the components were well interpretable. In Table 10.6 this solution is repeated. Only the loadings greater than .40 (in the absolute sense) are reported.

As has been explained in chapter 1, PCA on generalized correlation coefficients is rather limited in that it provides only loadings for the variables, without a representation for the objects (i.e., students) and for the categories of the variables. For this reason, the data have been

Table 10.6. *Loadings from the PCA solution reported by Vegelius and Bäckström.*

Variables	Components							
	1	2	3	4	5	6	7	8
1. Sex								
2. Year of birth								.67
3. Father's social group							.75	
4. Father's education							.75	
5. Mother's social group			.79					
6. Mother's education			.78					
7. Marital status								.68
8. Future degree					.81			
9. Future job					.82			
10. Father's religious observance	.62					.41		
11. Father's denomination	.85							
12. Father's participation in church worship						.69		
13. Mother's participation in church worship						.69		
14. Mother's religious observance	.62							
15. Mother's denomination	.82							
16. Student's denomination						-.43		.40
17. Student's participation in church worship		.70						
18. Student's praying		.65						
19. Student's participation in Communion		.64						
20. Student's reading the Bible		.53						
21. Political attitude								
22. Reaction to the Jesus Movement				.77				
23. Future of the Jesus Movement				.78				
24. Place of childhood and youth								

reanalyzed* here both by means of MCA and INDOQUAL. In contrast to Vegelius and Bäckström (1981) only two components have been retained in the present analyses. This choice of dimensionality was based partly on a comparison of the one–, two–, three–, and four–dimensional INDOQUAL solutions. The inertia accounted for by these solutions was 5.2 (6%), 8.3 (10%), 9.9 (12%), and 11.4 (14%), respectively. The percentages of inertia accounted for are very small, but, again it should be noted that in order to extract the most interesting information one may be satisfied with a solution that accounts only for a small part of the inertia of the data. The INDOQUAL solutions appeared to be nested approximately. For instance, the first two components in the three– and

* The author is obliged to Anders Bäckström and Jan Vegelius who kindly made the original data available.

four–dimensional solutions were closely related to the components in the two–dimensional solution. The third and fourth components were not very interesting in that they only pertained to two variables each. Obviously, the two–dimensional solution does not describe the data at hand exhaustively. However, it does seem to capture the most interesting relations between the variables in the enquiry. For this reason, in the sequel only the two–dimensional INDOQUAL solution is reported. For reasons of comparability the same dimensionality has been chosen for the MCA solution.

The (unrotated) loadings of the MCA solution are reported in Table 10.7.

Table 10.7. *Variable loadings resulting from MCA, MCA with varimax, and INDOQUAL.*

	MCA		MCA with varimax		INDOQUAL	
	1	2	1	2	1	2
1.	0.22	0.00	0.13	0.10	0.10	0.10
2.	0.04	0.22	0.08	0.18	0.04	0.10
3.	0.04	0.08	0.05	0.07	0.03	0.04
4.	0.10	0.28	0.24	0.14	0.15	0.06
5.	0.07	0.07	0.10	0.05	0.08	0.03
6.	0.04	0.10	0.11	0.03	0.05	0.02
7.	0.10	0.12	0.06	0.15	0.04	0.08
8.	0.31	0.17	0.04	**0.44**	0.02	**0.52**
9.	**0.48**	0.27	0.14	**0.62**	0.09	**0.67**
10.	**0.69**	0.20	**0.86**	0.04	**0.93**	0.02
11.	**0.70**	0.21	**0.85**	0.05	**0.93**	0.04
12.	**0.73**	**0.40**	**0.84**	0.28	**0.90**	0.10
13.	**0.71**	**0.40**	**0.82**	0.29	**0.87**	0.11
14.	**0.68**	0.16	**0.81**	0.04	**0.90**	0.02
15.	**0.62**	0.13	**0.70**	0.05	**0.80**	0.02
16.	0.06	0.13	0.06	0.14	0.05	0.11
17.	**0.60**	0.36	0.19	**0.76**	0.11	**0.86**
18.	**0.48**	0.36	0.08	**0.76**	0.04	**0.87**
19.	**0.54**	0.35	0.16	**0.73**	0.11	**0.86**
20.	**0.49**	0.29	0.22	**0.56**	0.14	**0.60**
21.	0.24	0.10	0.08	0.26	0.08	0.18
22.	0.19	0.01	0.13	0.06	0.09	0.07
23.	0.11	0.01	0.08	0.04	0.06	0.03
24.	0.08	0.05	0.06	0.07	0.05	0.04
sum	8.32	4.47	6.89	5.91	6.66	5.55
sum of squares	4.49	1.20	4.27	2.93	4.85	3.42

The MCA solution has been rotated to simple structure by means of varimax. The axes were rotated over an angle of 38°. The MCA loadings after rotation are reported in Table 10.7 as well. Finally, Table 10.7 contains the loadings resulting from the INDOQUAL solution, with loadings $\geq .4$ printed in bold face.

From Table 10.7 it is clear that the unrotated and rotated MCA solutions lead to different interpretations of the two components. That is, in the unrotated MCA solution many variables have rather high loadings ($\geq .4$), while on the second component only the variables 12 and 13 have high loadings. The first component can be interpreted as a rather general "Religious behavior of parents and student" component. The second component, on the other hand, can easily be interpreted as "Parent's participation in church worship", but this component does not capture much of the information in the data.

After varimax rotation the picture is quite different. Now only the variables 10, 11, 12, 13, 14, and 15 have high loadings on the first component, and the variables 8, 9, 17, 18, 19, and 20 have high loadings on the second component. The first component can be interpreted as "Religious behavior of parents", and the second component as "Student's religious behavior and plans". The components resulting from the rotated MCA solution seem to be much more interesting than those of the unrotated MCA solution.

The loadings of the two-dimensional INDOQUAL solution have been reported in Table 10.7 also. The components can be interpreted in much the same way as the components of the MCA solution after varimax rotation, because loadings greater than or equal to .4 are found in both analyses for the same variables. Nevertheless, the solutions differ systematically. It can be verified that loadings above .4 in the MCA solution correspond to higher loadings in the INDOQUAL solution, and, in all but one case, loadings below .4 in the MCA solution correspond to smaller loadings in the INDOQUAL solution. This clearly demonstrates that INDOQUAL tends to find more extreme loadings than MCA, even after varimax rotation. For the interpretation this does not make a lot of difference, however. The INDOQUAL components can be said to be interpretable in the same way as the rotated MCA components, but in the INDOQUAL solution this interpretation is brought out more clearly.

On the basis of the loadings alone, one can give only a rather global interpretation of the components by means of studying the loadings of the variables on the components. Because the variables are nominal variables, it

is of interest to see how these components can be interpreted in terms of the categories of the variables as well. In Table 10.8 the category coordinates (category centroids of object coordinates) from the INDOQUAL solution are given only for the variables with loadings greater than .4.

With these category coordinates the components of the INDOQUAL solution can be interpreted as follows. Component 1 (called "Religious behavior of parents" above), contrasts, for both parents, religious observance versus no religious observance, denomination to a certain church versus no answer, and, much versus little participation in church worship. Clearly, religious behavior can be further interpreted as "amount of participation in religious

Table 10.8. *INDOQUAL category coordinates of the variables with the highest loadings.*

	comp.1	comp.2
var.8: Future Degree		
Candidate of Theology	-0.09	-0.51
Bachelor of Arts	0.31	1.39
Other degree	-0.04	-0.16
Don't know	-0.06	-0.08
var.9: Future Occupation		
Clergyman	-0.14	-0.51
Teacher	0.38	1.43
Deacon	0.52	-0.55
Other occupation	-0.45	-0.40
Don't know	-0.14	-0.35
var.10: Father's religious observance		
Yes	-1.08	-0.16
No	0.85	0.10
Don't know	0.86	0.27
var.11: Father's denomination		
Church of Sweden	-1.07	-0.29
Other denomination	-1.13	0.21
No answer	0.85	0.12
var.12: Father's participation in worship		
Never	0.91	0.46
At least once a year	0.83	0.23
At least six times a year	0.60	-0.53
At least once a month	-1.00	-0.20
At least once a week	-1.14	-0.12
Don't know	0.82	-0.45

Table 10.8. *INDOQUAL category coordinates of the variables with the highest loadings* (*continued*).

	comp.1	comp.2
var.13: Mother's participation in worship		
Never	0.89	0.63
At least once a year	0.90	0.21
At least six times a year	0.72	-0.46
At least once a month	-0.94	-0.18
At least once a week	-1.13	-0.13
Don't know	0.48	0.47
var.14: Mother's religious observance		
Yes	-1.00	-0.15
No	0.90	0.11
Don't know	0.89	0.35
var.15: Mother's denomination		
Church of Sweden	-0.85	-0.17
Other denomination	-1.02	0.01
No answer	0.89	0.13
var.17: Student's participation in worship		
Not for years	0.37	2.35
At least once a year	0.65	1.95
At least six times a year	0.52	0.70
At least once a month	0.16	-0.47
At least once a week	-0.28	-0.44
var.18: Student's praying		
Today or yesterday	-0.10	-0.39
One week ago at most	0.38	-0.42
One month ago at most	0.07	2.15
Not for years	0.46	2.19
Don't know	0.63	2.29
var.19: Student's participation in Communion		
One week ago	-0.19	-0.47
One month ago	-0.02	-0.37
Six months ago	0.72	-0.47
One year ago	-1.20	0.15
Not for years	0.42	2.03
Don't know	0.88	0.36
var.20: Student's reading the Bible		
Today or yesterday	-0.23	-0.46
One week ago	0.27	-0.04
One month ago	0.51	0.79
Six months ago	-0.89	2.71
Not for years	0.74	1.84
Don't know	1.07	-0.97

activities". On the basis of this interpretation one might suspect that the category "No answer" of variables 11, and 15, in fact means "No member of any denomination". The second component (called "Student's religious behavior and plans" above) can be interpreted as the contrast between students preparing for a clerical profession versus a profession as a layman (variables 8 and 9), as well as the "amount of the student's participation in religious activities" (variables 17–20). Again, interpretation on the basis of category coordinates complements the previous interpretation in terms of the variables in a useful way.

Finally, the results reported here are compared to those of Vegelius and Bäckström (1981). It can be seen that the components 1 and 6 in the latter solution are practically completely comprised in the first component in the INDOQUAL solution, and that the components 2 and 5 in the Vegelius and Bäckström solution are summarized well by the second INDOQUAL component. This indicates that the Vegelius and Bäckström solution provides components that could well be combined into fewer components. On the other hand, the third, fourth, seventh and eighth components in the Vegelius and Bäckström solution do not reappear in the INDOQUAL solution. This may be a consequence of the choice of a rather small dimensionality. In fact, the third component of the three–dimensional INDOQUAL solution is closely related to the seventh component in the Vegelius and Bäckström solution, and slightly related to their third component as well. These and higher components have not been considered in the INDOQUAL solution reported here.

10.4. The abortion survey: Components analysis of mixed variables

Gifi (1981, pp. 357–360) has described data from a survey among 575 respondents on attitudes with respect to abortion, capital punishment, euthanasia, and sexual freedom. In addition, the scores on several background variables were available. These data have been analyzed by Gifi by means of several options of the PRINCALS program. In the present section part of these data* is analyzed by means of INDOMIX, and the solution will be compared to that of PCAMIX.

* The author is obliged to Jacqueline Meulman who kindly made the data available.

The variables on capital punishment are hardly related to the other variables, as was found by Gifi (1981), and have therefore been excluded from the analysis, just as has been done by Gifi. In contrast to what has been done by Gifi, however, no other variables have been left out of the analysis. The data set contained many missing data. Because the present version of the INDOMIX program does not allow for missing data, all respondents with one or more missing data (22 %) have been deleted. The remaining sample contained 446 respondents. Of the 34 variables that remained after deletion of the capital punishment variables, two variables have been recoded as follows: the eleven categories of the variable "Present profession or job" (FUN) have been reduced to nine categories by merging the categories "managerial, more than ten employees" and "managerial, less than ten employees", and the categories "free profession" and "independent farmer"; the seven categories of the variable "degree of urbanization" (URB) have been reduced to five categories by merging the three categories "Amsterdam", "Rotterdam", and "The Hague". The other variables have been analyzed in their original form, as described by Gifi (1981). They are labelled here (as in Gifi) as A1 through A16 (questions concerning abortion), EU1 through EU5 (questions concerning euthanasia), SF1 through SF5 (questions concerning sexual freedom), SEX (male, female), AGE (six age categories), SOC (eight social class levels), REL (religion: four categories), POL (political preference: five categories), and EDU (four levels of education). Together with FUN and URB this leads to the total of 34 variables analyzed here.

The variables are rather different in type. The variables A1 through A8 and EU1 through EU5 are binary variables (agree or disagree with a statement). The variables A9 and A10 form six–point scales about the duration of pregnancy after which abortion is justifiable. The categories range from "until three months" (1) to "after six months" (5), while category 6 stands for "not justifiable". The latter category has been recoded as 0. The variables A11 through A14, and SF1 through SF5 are five–point Likert scale items. These variables, as well as the variables A9 and A10, have been considered as quantitative variables, because they clearly implied an ordinal scale, and, according to the theory in chapter 3, one way of dealing with ordinal variables is to treat them just as quantitative variables. The background variable SOC is also treated as a quantitative variable, because its eight categories are clearly ordered. Variables A15 and A16, just as the background

variables REL, POL, EDU, FUN, and URB, have been been considered as nominal variables.

The data described above have been analyzed by both PCAMIX and INDOMIX. The variables have not been standardized, but some of the variables have been weighted. That is, in order to let the background variables serve

Table 10.9. *Loadings from PCAMIX and INDOMIX.*

	PCAMIX		INDOMIX		INDOMIX loadings for quantitative variables	
	comp.1	comp.2	comp.1	comp.2	comp.1	comp.2
A1	0.00	0.19	0.02	0.04		
A2	**0.63**	0.00	**0.63**	0.00	0.79	
A3	**0.26**	0.10	**0.29**	0.00	0.54	
A4	**0.59**	0.00	**0.60**	0.00	0.78	
A5	0.19	0.13	**0.23**	0.02	0.48	
A6	**0.61**	0.00	**0.63**	0.00	0.80	
A7	**0.66**	0.00	**0.67**	0.00	0.82	
A8	**0.57**	0.00	**0.57**	0.00	0.75	
A9	**0.42**	0.03	**0.43**	0.01	-0.66	
A1	**0.48**	0.02	**0.49**	0.01	-0.70	
A1	**0.59**	0.01	**0.60**	0.00	0.77	
A12	**0.37**	0.18	**0.48**	0.03	-0.69	
A13	**0.49**	0.13	**0.61**	0.02	-0.78	
A14	**0.35**	**0.20**	**0.47**	0.03	-0.69	
A15	**0.33**	0.03	**0.26**	0.00		
A16	**0.21**	0.05	**0.20**	0.01		
EU1	**0.27**	0.07	**0.26**	0.00	0.51	
EU2	**0.22**	0.04	**0.20**	0.00	0.45	
EU3	**0.31**	0.05	**0.30**	0.00	0.55	
EU4	**0.21**	0.15	**0.23**	0.02	0.48	
EU5	0.10	0.00	0.07	0.00		
SF1	0.04	0.15	0.06	0.09		
SF2	0.02	**0.34**	0.07	**0.22**		0.47
SF3	0.00	**0.45**	0.02	**0.76**		0.87
SF4	0.06	**0.37**	0.12	**0.22**		0.47
SF5	0.02	**0.56**	0.06	**0.68**		0.83
SEX	0.00	0.00	0.00	0.00		
AGE	0.01	0.14	0.02	0.14		
SOC	0.00	0.01	0.00	0.01		
REL	**0.21**	0.06	**0.24**	0.04		
POL	**0.22**	0.11	**0.27**	0.02		
EDU	0.01	0.07	0.01	0.07		
FUN	0.03	0.07	0.04	0.05		
URB	0.05	0.00	0.04	0.01		

mainly as "supplementary" or "passive" variables that do not actually affect the solution, these variables have been given a very small weight (0.001) in the analyses.

In all analyses the dimensionality of the solutions was set at two, because the first two components of the three-dimensional INDOMIX solution hardly differed from those of the two-dimensional solution, and the third component essentially represented only one variable. The PCAMIX solution accounted for 12% of the inertia by means of the SUMPCA model ($IAF_S = .12$), while it accounted for 17% of the inertia in terms of the INDORT model ($IAF_I = .17$). The INDOMIX solution accounted for 19% of the inertia ($IAF_I = .19$). Clearly, the INDORT model fits the data considerably better than the more restricted SUMPCA model.

The loadings from PCAMIX (after varimax rotation) and INDOMIX are given in the first four columns of Table 10.9. Loadings $\geq .2$ have been printed in bold face. Clearly, the loadings from PCAMIX and INDOMIX differ, but when the components are interpreted on the basis of the variables that load high on them, the components are interpreted almost identically in the two analyses. That is, the variables A2 through A16 and EU1 through EU4 have high loadings on the first component, and the variables SF2 through SF5 have high loadings on the second component. In addition, the first component is related to the background variables REL and POL.

For a more detailed interpretation, it is useful to note that for quantitative variables (including binary variables) the loadings given here are the squares of the product-moment correlations. For the quantitative variables with INDOMIX loadings higher than .2 the ordinary product-moment correlations between the variables and the components concerned, are given in the last two columns of Table 10.9. The interpretation of the components in terms of the nominal variables can be facilitated by inspection of the category coordinates. For the variables A15, A16, REL and POL these are given in Table 10.10.

On the basis of Tables 10.9 and 10.10 the first INDOMIX component can be interpreted as "liberal versus conservative". This can be seen as follows. The first component is strongly positively correlated with the variables A2 through A8. These items ask whether (1) or not (2) abortion is allowed under specified circumstances. This component is negatively correlated with A9 and A10, asking after how much time abortion is still justifiable (from "not

Table 10.10. *Category coordinates on the INDOMIX dimensions for four high loading nominal variables.*

	comp.1	comp.2
A 15: Should abortion requests be handled by law ?		
law desired allowing for abortion only in special cases	0.50	0.01
law desired making abortion difficult	0.23	-0.19
no law desired, doctor decides in abortion requests	-0.57	0.03
A 16: Should abortion be permitted after twelve weeks of pregnancy ?		
after 12 weeks absolutely forbidden	0.65	-0.12
abortion after 12 weeks in special cases only	-0.11	0.02
law should not specify a time limit	-0.60	0.13
REL: Religion		
Reformed	0.24	-0.43
Calvinist	0.76	0.18
Roman Catholic	0.30	0.04
none	-0.57	0.10
POL: Political preference		
left	-0.38	-0.04
denominational	0.70	-0.02
liberal	-0.41	0.19
right	1.49	-0.68

justifiable" to after "six months"). Furthermore, the variables A11 to A14 correlate strongly with the first component. These variables indicate on five–points scales whether the respondents agree (low scores) or disagree (high scores) with certain moralistic statements on abortion. The first of these (A11) is a pro–abortion statement, which correlates positively with the first component, the others (A12 through A14) are against abortion, and correlate negatively with the first component. The variables EU1 through EU4 also correlate (positively) with the first component. These questions ask whether (1) or not (2) euthanasia is justifiable under certain specified circumstances. Summarizing, one can see that persons with low scores on the first component typically allow for abortion in many circumstances, that is, agree with many of the statements A2 through A8, find abortion justifiable even after long periods of pregnancy (A9 and A10), agree with the pro abortion statement (A11) and do not agree with the contra abortion statements (A12 through A14), find that laws regulating abortion or setting time limits regarding abortion are not needed (A15 and A16), and find that euthanasia is

justifiable under several circumstances (EU1 through EU4), while people with high scores typically take the reversed stand. This explains why this component can be labeled as "liberal versus conservative". As has been said earlier, two background variables are also related to this component. Upon inspection of the category coordinates of these variables (REL and POL), it is clear that at the lower end of this dimension one finds people without religion, and with left or liberal political preference. At the other end of the dimension one finds Reformed, Calvinist, and Roman Catholic respondents, with denominational or conservative political preferences.

The second component is positively correlated with four variables, SF2 through SF5. These variables are five–point ratings (from agree completely (1) to disagree completely (5)) of statements that are clearly contra sexual freedom. This component can hence simply be interpreted as "contra versus pro sexual freedom". This completes the description of the analysis of the abortion survey data.

As has been mentioned in section 10.1.2, one way of studying the stability of one's data is by means of cross–validation. This can be done as follows. The data set is split into two halves, which are both reanalyzed. These analyses yield weights (for the categories and the quantitative variables) which can be applied to the other half of the data (see section 10.1.2). The resulting components of object coordinates are not necessarily uncorrelated, but apart from this one can handle them as ordinary INDOMIX components. In order to see how sensitive the analysis is to particular characteristics of the sample drawn, the component scores and the loadings that can be computed after applying the weights to one half of the data are compared to the component scores and loadings resulting from INDOMIX on this half of the data. When the component scores and the loadings are very much the same, one can conclude that the analysis is hardly sensitive to the data. The component scores from the two analyses are compared by computing the correlation between them. In order to compare the loadings from one analysis to that of another, we use the coefficient proposed by Gower (1971) for the association between two numerical variables. In our case, this coefficient is equal to one minus the mean of the absolute differences between the loadings from the different solutions on a component. It has a maximum of one, which is attained when the loadings are identical.

For the present data this kind of cross–validation seems more useful than

a (random) jackknife study. Jackknifing is meant for studying changes under deletion of just one individual, which seems most appropriate in samples that intend to capture almost the whole population (as in the cetacea data). When a sample is intended to represent a much larger population, the present cross–validation procedure seems more appropriate, because it is based on comparing subsamples that should, just as the original sample, be representative for the population.

The cross–validation study undertaken here started with splitting the sample into two groups by alternatingly assigning one individual to the one and the next to the other group. First, the weights resulting from the INDOMIX analysis of the second half have been applied to the first half. The resulting components were hardly correlated (p.m.c. equal to –.13). They correlated strongly with the components resulting from INDOMIX on this half of the data, that is, the correlation between the first components was .997 and that between the second components was .978. The resulting loadings have been compared with the loadings from an INDOMIX analysis on the first half, by computing Gower's coefficients over the loadings on the (corresponding) components of the different solutions. This coefficient was .991 for the first components and .988 for the second component. These values correspond to mean absolute differences between the loadings of 0.009, and 0.012, respectively.

The inverse procedure of applying weights resulting from the first half to the data in the second half resulted in components that were nearly uncorrelated (p.m.c. equal to .10). Again, they correlated strongly with the components resulting from INDOMIX on this half of the data, that is, .997 for the first components and .980 for the second components. The coefficients for comparing the resulting loadings were 0.991 for the first component, and 0.987 for the second component, corresponding to mean absolute differences between the loadings of 0.009, and 0.013, respectively.

Clearly, the correlations between corresponding components and the values of Gower's coefficient for comparison of the loadings on the corresponding components are all very high. On the basis of this cross–validation study it can be concluded that the INDOMIX solution reported above is not very sensitive to sample fluctuations.

10.5. Residual complaints after head injury: Components analysis of binary variables

Van Zomeren and Van den Burg (1985) have observed a number of patients who had incurred a severe closed head injury. Many of them still reported some residual complaints, even after two years after the accident. These complaints have been recorded by means of 17 items, listed in Table 10.11. Apart from these (binary) variables, two measures indicating the severity of the injury have been scored as well. These latter variables are post–traumatic amnesia (PTA), and the extent to which previous work could be resumed (RTW).

The data on 50 patients have been reanalyzed* in the present study. Two of the original 52 patients have been excluded from the analysis because some observations were missing on them. The variables PTA and RTW are ordinal variables. They have been dichotomized here for simplification. PTA could very well be dichotomized by using the cut–off point proposed by Van Zomeren and Van den Burg (at 13 days). RTW was a five point–scale which could be distinguished roughly into "former work resumed" versus "former work resumed only partly or not at all". This variable has been dichotomized accordingly, yielding a set of 19 dichotomous variables.

The data, transformed as described above, have been analyzed by both MCA and INDOQUAL. Because the variables are all dichotomous, MCA comes down to PCA, and instead of INDOQUAL, INDOMIX can be used while all variables are considered numerical, which is computationally attractive (as has been discussed in chapter 9). The MCA analysis is followed by a varimax rotation.

In both analyses, two–dimensional solutions have been studied. This dimensionality has been chosen partly for reasons of comparison with the PCA solution reported by Van Zomeren and Van den Burg, and partly because it yielded an interpretable solution, in contrast to, for instance, the three–dimensional INDOQUAL solution for these data. In contrast to the other examples in the present study, the loadings that are reported here are ordinary (point–biserial) correlations between the variables and the

* The author is obliged to Pim van den Burg for kindly providing the original data

components, just as in ordinary PCA. These loadings are recorded in Table 10.11. For reasons of comparison, the last two columns of Table 10.11 contain the PCA–loadings reported by Van Zomeren and Van den Burg (1985). Loadings $\geq$ .60 are printed in bold face.

Table 10.11. *Loadings resulting from MCA and INDOQUAL, as well as loadings reported by Van Zomeren and Van den Burg* (1985).

	MCA + varimax		INDOQUAL		Original loadings	
	comp.1	comp.2	comp.1	comp.2	comp.1	comp.2
PTA	**0.76**	-0.09	**0.74**	0.00	-0.13	**0.80**
RTW	**0.71**	0.10	**0.73**	0.10	0.14	**0.70**
forgetfulness	**0.65**	0.19	**0.71**	0.07	0.19	**0.63**
irritability	0.13	0.54	0.21	0.37	0.59	-0.03
slowness	**0.69**	0.06	**0.61**	0.16	0.25	**0.66**
poor concentration	0.52	0.53	**0.63**	0.30	**0.61**	0.42
fatigue	0.39	**0.64**	0.42	0.53	**0.68**	0.31
dizziness	0.14	0.44	0.17	0.32	0.52	-0.03
incr.need of sleep	0.00	0.37	0.12	0.04	0.51	-0.24
intol. of light	0.10	**0.68**	0.25	0.30	**0.72**	-0.07
intol. of noise	0.42	0.52	0.30	**0.83**	**0.61**	0.33
loss of initiative	**0.65**	0.21	**0.72**	0.06	0.51	0.38
headache	0.13	**0.64**	0.07	0.17	0.57	-0.17
crying more readily	0.32	0.46	0.37	0.23	**0.60**	0.16
inability to do two things simultan.	**0.68**	0.38	**0.66**	0.42	0.44	**0.62**
intol. of bustle	0.22	0.55	0.09	**0.93**	**0.61**	0.12
depressed mood	**0.69**	0.04	**0.70**	0.01	0.32	0.53
more anxious	0.00	0.40	0.09	0.14	0.33	-0.05
indifference	0.04	0.48	0.09	0.18	0.24	0.13

The differences between the PCA loadings from Van Zomeren and Van den Burg (1985) and the MCA loadings reported here can be attributed to several differences in the procedures followed. Van Zomeren and Van den Burg (1985) did not dichotomize PTA and RTW, and performed a quartimax rotation instead of a varimax rotation. Especially the latter may well explain the differences,

because a quartimax rotation tends to yield a strong general component, which corresponds to the finding that Van Zomeren and Van den Burg's first component can be considered quite general. In order to check this, the solution reported by Van Zomeren and Van den Burg has been matched to the MCA solution, via an orthogonal Procrustes rotation (Green, 1952). The rotated components were related highly to the MCA components. That is, the congruence coefficient (Tucker's ϕ) measured between the loadings on the first components was .99, and that between the second components .96, indicating that the loadings yield equivalent interpretations.

The INDOQUAL solution differs from the MCA solution especially with respect to the second components (Tucker's ϕ equal to .84). The Procrustes rotation did not improve this. In MCA this component seems hard to interpret, because the variables that have high loadings on it do not seem to have much in common, except some forms of intolerance. In INDOQUAL the component can be seen as a dimension expressing more specifically intolerance of noise and bustle. In both analyses, the first component can be interpreted in the same way as Van Zomeren and Van den Burg (1985) did for their second component, that is, as a "severity"–dimension. Just as in Van Zomeren and Van den Burg (1985), this component is highly related to PTA and RTW, as well as to forgetfulness, slowness, poor concentration, inability to do two things simultaneously, and depressed mood. Unlike in Van Zomeren and Van den Burg (1985), in both analyses the first component is also related to loss of initiative, which corresponds to what one might have expected.

10.6. Characteristics of alcoholic and nonalcoholic drinks: Effects of standardizing nominal variables

For testing the INDOMIX program the author has created some fictitious data sets that are based on common sense knowledge. The data set to be considered in the present section has been constructed as follows. Some characteristics of 34 drinks, both soft–drinks and alcoholic drinks, have been given in terms of the following (seven) variables: alcoholic strength (five categories, from no alcohol to over 30 %), addition of sugar (yes or no), does drink contain carbonic acid (yes or no), kind of raw product which essentially determines the taste of the drink (fruit, grain, herbs, artificial flavors), price (four categories from cheap to expensive), taste (very sweet, sweet,

dry, bitter), and color (colorless, red, light–red, yellow, brown). The scores of the drinks on the variables are fictitious in that they are based on the author's (limited) knowledge of these drinks. As a consequence, certain data might actually be technically incorrect, but in general it can be expected that these data reflect reality rather well. The data with the names of the drinks (some of which are exclusive for certain European countries) are given in Table 10.12.

Table 10.12. *Characteristics of* 34 *drinks.*

variables	Alcohol	Sugar	Carbon	Raw prod.	Price	Taste	Color
	5 cat	2 cat	2 cat	4 cat	4 cat	4 cat	5 cat
labels							
syrop	1	1	2	4	2	1	4
cola	1	1	1	4	1	1	5
seven–up	1	1	1	4	1	1	1
orangina	1	1	1	1	1	1	4
apple juice	1	2	2	1	1	1	4
orange juice	1	2	2	1	2	2	4
red bordeaux	2	2	2	1	2	3	2
wh. bordeaux	2	2	2	1	2	3	4
red Lambrusco	2	2	1	1	2	1	2
rosé	2	2	2	1	2	2	3
Moselle wine	2	2	2	1	2	1	4
Sekt	2	2	1	1	2	2	4
Riesling	2	2	1	1	2	3	4
champagne ds	2	2	1	1	4	2	4
champagne br	2	2	1	1	4	3	4
sherry	3	2	2	1	3	2	5
port	3	2	2	1	3	1	2
Cointreau	5	1	2	1	4	1	1
jenever	5	2	2	2	3	4	1
gin	5	2	2	2	4	4	1
whisky	5	2	2	2	4	4	4
beer	2	2	1	2	1	4	4
old–br. beer	2	1	1	2	2	4	5
guinness	2	2	1	2	2	4	5
cider	2	1	1	1	2	1	4
strawberry lq	4	1	2	4	3	1	3
banana liquor	4	1	2	4	3	1	4
cherry brandy	4	1	2	4	3	1	3
bl.currant lq	4	1	2	4	3	1	2
slivovic	5	2	2	1	4	4	1
ouzo	5	1	2	3	4	1	1
Pernod	5	1	2	3	4	1	1
Jägermeister	4	2	2	3	3	4	5
rum	5	1	2	2	4	2	1

These data have been analyzed by means of INDOQUAL with and without standardization of the variables. As has been explained in chapter 3, standardization of nominal variables comes down to weighting these variables by $(m_j-1)^{-1/2}$. Clearly, standardizing the variables in the present data set implies using much larger weights for the variables Sugar and Carbon than for the other variables. In section 4.4 some grounds for choosing whether or not to normalize one's variables have been discussed. In the present section the effects of these choices on the results of an INDOQUAL analysis of the data set described just above will be compared.

The dimensionality of the solutions has been set, rather arbitrarily, to two after verifying that it gave an interpretable solution. Table 10.13 reports the loadings of the seven variables on the 2 components in both solutions. Loadings $\geq .70$ are printed in bold face.

Table 10.13. *Variable loadings resulting from INDOQUAL with and without standardization, respectively.*

	INDOQUAL on standardized variables		INDOQUAL on nonstandardized variables	
	comp.1	comp.2	comp.1	comp.2
Alcohol	0.38	0.56	**0.80**	**0.94**
Sugar	**0.96**	0.00	0.57	0.02
Carbon	0.00	**0.98**	0.06	0.10
Raw product	0.57	0.08	**0.84**	0.33
Price	0.11	0.42	0.42	**0.71**
Taste	0.53	0.02	0.54	0.12
Color	0.10	0.22	0.13	0.83

As is clear from Table 10.13, the INDOQUAL solutions on standardized and nonstandardized variables differ markedly. When the data are standardized, the solution is, in fact, dominated by the two binary variables Sugar and Carbon, the variables receiving the largest weights due to the standardization. Some of the other variables have modest loadings on these components as well, but by no means as high as those of the binary variables. In the INDOQUAL analysis of nonstandardized variables the binary variables

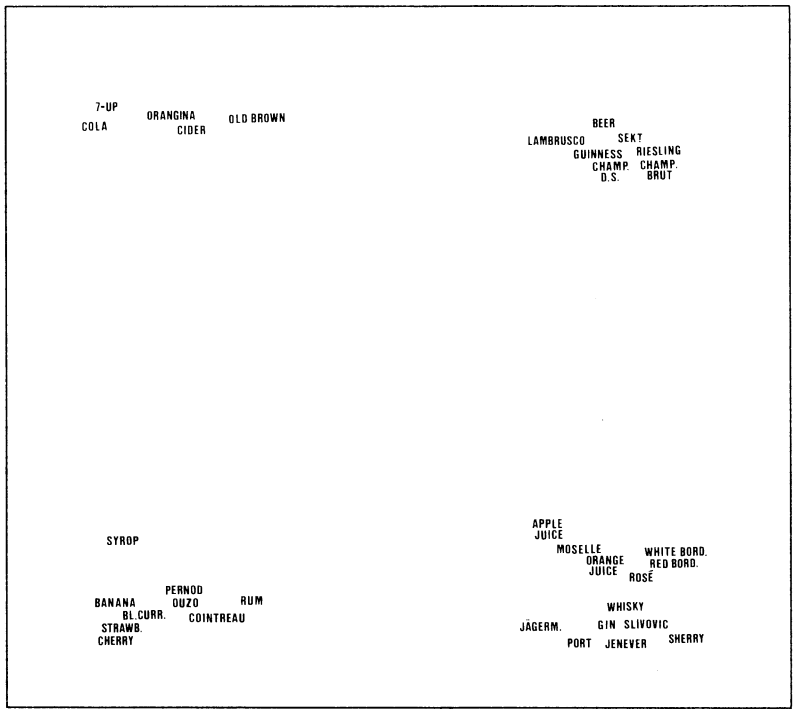

Figure 10.4. *Object coordinates resulting from INDOQUAL on standardized variables.*

only play modest roles. This solution is based more on the amount of alcohol in the drinks, the raw product determining the taste, the price, and the color of the drinks. The latter solution is more informative than the first in that it represents the data in terms of more variables.

The object coordinates of the two solutions are given in Figures 10.4 and 10.5 for the INDOQUAL solution of standardized variables and of nonstandardized variables, respectively. In the analysis of the standardized variables, the drinks are clustered in four groups with characteristics "sugar and CO_2 added", "contains CO_2 but no sugar added", "sugar added but no CO_2", "neither sugar nor CO_2 added". This results, for instance, in a cluster containing fruit juices, several wines, and strong drinks like gin and whisky at the same time, which does not seem to correspond to how one would typically classify drinks. In the solution of INDOQUAL on nonstandardized variables the drinks are not clustered as clearly as in the other analysis, but the

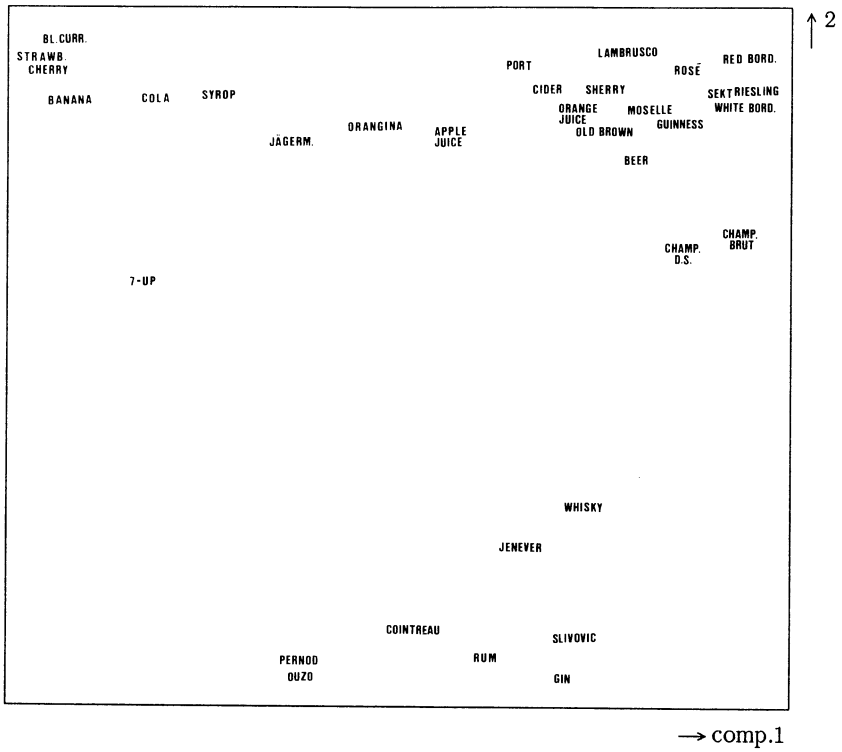

Figure 10.5. *Object coordinates resulting from INDOQUAL on nonstandardized variables.*

picture does seem to make more intuitive sense, especially separating strong drinks from soft drinks.

It can be concluded that, at least in the present case, it is not useful to standardize the qualitative variables. The original variables can be considered to yield quantification matrices in comparable units, and what is called standardizing the variables might be seen as, in fact, un–standardizing the variables, that is, letting variables with few categories have the strongest impact on the solution.

10.7. Italian freight transportation data: A comparison of INDORT and TUCKALS-3 on quantification matrices

Marchetti (1988) has described a data set collected by the Italian National Research Organization (CNR) on eight characteristics of 54 Italian

freight transportation industries. These characteristics are expressed in terms of eight variables, which are considered as nominal variables both here and by Marchetti (1988). These variables are labelled A through H, just as has been done by Marchetti (1988). They measure:

A: number of employees (three categories)

B: number of tractors (three categories)

C: use of containers (binary: yes or no)

D: number of semi–trailers (four categories)

E: number of transports in 1981 (three categories)

F: juridical status (four categories: joint–stock company, limited liability company, general or limited partnership, sole traders)

G: location (three categories: north, center, south)

H: type of firm (four categories: shipping agent, forwarding agent and carrier, lorry–conveyor, carrier).

Variables A, B, D and E are polytomized quantitative variables, but the ordering underlying the categories is not used here, following Marchetti (1988).

Marchetti (1988) has analyzed these data by means of TUCKALS–3 applied to quantification matrices. As quantification matrices he has chosen the standardized versions of the ones used in MCA and INDOQUAL. In other words, he has used the same quantification matrices as are used in MCA and INDOQUAL, but applies weights to the variables in order to standardize them. These weights are $(m_j–1)^{-½}$, $j = 1,...,m$, where m_j is the number of categories of variable j. As has been noted by Marchetti and has been seen in section 10.6, this may result in a solution in which variables with a small number of categories are better represented than the others.

In the present study the same data are analyzed by means of INDOQUAL, and the resulting loadings and category coordinates are compared to the ones found by Marchetti (1988). For the TUCKALS–3 analysis Marchetti (1988) has chosen both dimensionalities (i.e., for the components of variables and for the components of objects) equal to three. This corresponds best to choosing the dimensionality of the INDOQUAL solution equal to three as well.

The three–dimensional INDOQUAL solution yields a function value of 2.43, which can be seen as the inertia accounted for. Because the variables are standardized, the total inertia in the data is equal to the number of variables, that is, 8. As a result, INDOQUAL accounts for 2.43*100/8 = 30.4 %

of the total inertia. Marchetti reports that the TUCKALS-3 solution accounted for 30.8 % of the inertia. It can be concluded that INDOQUAL represents the information in the data almost as well as TUCKALS-3 does, while INDOQUAL uses a much simpler model.

The loadings for the variables found by the INDOQUAL solution are given in Table 10.14. These loadings were computed as $\mathbf{x}_l' S_j \mathbf{x}_l$, $j = 1,...,8$, $l = 1,...,3$, where S_j is the *standardized* version of the quantification matrix used in MCA and INDOQUAL. Loadings $\geq .50$ are printed in bold face.

Table 10.14. *Variable loadings resulting from INDOQUAL.*

	comp.1	comp.2	comp.3
A	**0.55**	0.02	**0.70**
B	0.13	0.10	0.02
C	0.00	**0.99**	0.00
D	0.23	0.26	0.18
E	**0.56**	0.05	0.02
F	0.24	0.05	0.03
G	0.21	0.03	0.06
H	0.21	0.12	0.00

It can be seen that only the variables A, C, and E are represented quite well, which is the same conclusion as was drawn by Marchetti (1988). This is about the only conclusion made by Marchetti as far as the variables are concerned. From his plots for the variables one might conclude that the first component mainly represents variables A and C, and to a less extent also D and E. The second component contrasts C to the other variables, and the third component contrasts A to E, while the other variables are in between these extremes. It seems difficult to give a further interpretation of these results on the basis of the loadings alone, especially as far as contrasts between nominal variables are concerned. A simple structure rotation might possibly have helped here.

The loadings from the INDOQUAL solution lead to a different interpretation. The first component of the INDOQUAL solution is highly related to variables A and E, and can be interpreted as a component expressing the "size of the company". The second component only represents variable C well, and the third component mainly represents variable A. A more detailed

comparison of the loadings found here and by Marchetti is not feasible because Marchetti (1988) only gave plots of the variables. In addition, it should be noted that the TUCKALS–3 solution can be rotated by any nonsingular matrix, hence comparing the results of the two analyses should take such rotational freedom into account. From the plots, however, it is practically impossible to see how such a rotation should be made.

Apart from plotting the loadings for the variables, Marchetti (1988) also gives a plot of the category coordinates for the first two components, and he provides a rather detailed interpretation of these results. This plot is reproduced here (with permission from Marchetti) in Figure 10.6. The capitals denote the variables, and the indices denote the categories of the variables concerned. In the same plot the category coordinates for the first two INDOQUAL components have been given after a rotation (by hand) to maximal agreement of the two configurations. The INDOQUAL category coordinates are given by lower case characters. The INDOQUAL coordinate axes are also depicted in this plot, and labelled as "DIM.1" and "DIM.2",

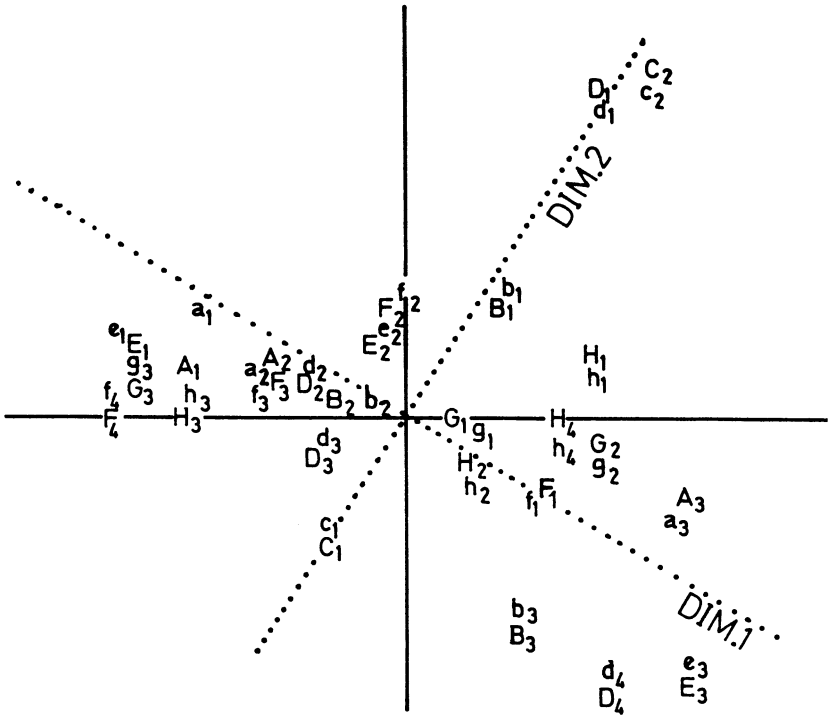

Figure 10.6. *Plot of the category coordinates on the dimensions* 1 *and* 2 *from the TUCKALS (capitalized) and INDOQUAL (lower case) solutions.*

respectively. Clearly, after rotation, the configurations of category points resulting from the two different analyses are virtually equal, and the interpretations provided by Marchetti hold for the INDOQUAL solution as well.

The components found by TUCKALS–3 for the objects and for the variables are related to each other via the elements of the core matrix. In this way TUCKALS–3 represents the data by means of a more complicated model than INDOQUAL does. In the latter method the components for the objects and those for the variables have a simple one–to–one relation with each other. Therefore, it is possible to interpret the loadings by relating them to the category coordinates and vice versa. In the present example such an interpretation can be used to interpret the first INDOQUAL component as expressing the size of a company (variables A and E mainly), and the second component as the one that contrasts the use of containers to the use of semi–trailers (variables C and to a less extent D). On the basis of the TUCKALS–3 solution a similar interpretation can be made, based on the category coordinates alone, but this interpretation of the components cannot readily be carried over to the components of the variables. Because the INDOQUAL solution accounts for practically the same amount of inertia as the TUCKALS–3 solution does, it seems that the former is to be preferred for the representation of the present data.

10.8. The Sugiyama Data: Where INDOQUAL fails

INDOQUAL has been developed as an alternative to MCA in order to find better representations for data in which subsets of variables form clusters of closely related variables, while variables of different clusters are hardly related. Such data are often characterized by rather large (generalized) correlations within subsets of variables. However, if nominal variables do not have high generalized correlations among each other, the data set may nevertheless contain interesting information. This can be the case, for instance, with preference data, e.g., binary variables indicating whether or not a stimulus is picked out of a number of stimuli. The analysis of such data has been described, for instance, by Heiser (1981). He gives several example data sets, one of which, the "Sugiyama data" (Sugiyama, 1975, see Heiser, 1981, p.142) is reanalyzed here. The data consists of six binary variables

pertaining to religious behavior. The variables can be described briefly as A: Do you make it a rule to practice religious conduct; B: Do you visit a grave once or twice a year; C: Do you occasionally read religious books; D: Do you visit shrines and temples to pray; E: Do you keep a talisman; and F: Did you draw a fortune. For the exact wordings of the questions, as well as the data themselves, the reader is referred to Heiser (1981, p.142).

Heiser has analyzed these data by MCA and found no interpretable results. A different method, better adapted to the analysis of binary proximity data (called HOMANA–BIN by Heiser, 1981), did yield good results. It turned out that the questions could be seen to form a scale ordered as F–D–E–B–A–C. In the present study the data have been analyzed by means of INDOQUAL. It turned out, however, that INDOQUAL did not produce a useful representation for this data. That is, INDOQUAL consistently yielded solutions of which each component represented one variable almost exclusively. Using different random starts tended to yield different solutions with almost the same function value, but with quite different loadings.

Clearly, these INDOQUAL results are of no value whatsoever. As has been said above, INDOQUAL is supposed to be useful when the data contains some subsets of highly correlated variables. Such subsets might be identified by inspection of the correlations between the variables, possibly complemented by a PCA of these correlations. In Table 10.15 the ϕ^2–contingency coefficients are given for this data. Clearly, no subsets of highly correlated variables are present in this data. This explains the poor quality of the INDOQUAL solutions.

Table 10.15. *ϕ^2 – contingency coefficients among the variables of the Sugiyama data.*

	A	B	C	D	E	F
A	1.00					
B	0.01	1.00				
C	0.08	0.00	1.00			
D	0.01	0.01	0.00	1.00		
E	0.01	0.03	0.00	0.08	1.00	
F	0.00	0.00	0.00	0.04	0.04	1.00

10.9. Concluding remarks

In the present chapter seven example data sets have been analyzed by some of the techniques described in the present study. There are considerable differences between these examples, as well as between the choices that have been made for analyzing them. Nevertheless, some general remarks can be made.

In most of the analyses reported here PCAMIX solutions have been compared with INDOMIX solutions. Except for the example in section 10.5, the INDOMIX solutions did not differ much from the PCAMIX solutions, at least, after the latter had been rotated by means of varimax. The modest differences that could be observed, however, did reveal a systematic tendency. High loadings in PCAMIX correspond to even higher loadings in INDOMIX, and small loadings in PCAMIX correspond to even smaller loadings in INDOMIX. This has been pointed out in particular in sections 10.2 and 10.3, but can be found in other examples as well, although not always as clear.

In two cases the stability of the INDOMIX solution has been studied. In both cases the solution appeared highly stable. Obviously, this is only partly a feature of the method. Stability is first of all determined by the homogeneity of a population or the representativeness of a sample. The special choices made in the analysis may also affect the stability. For instance, the dimensionality of the solution might be related to its stability. In the present analyses mainly small dimensionalities have been chosen. These choices might partly account for the stability of the solutions. Implicitly, this reasoning gives another criterion for determining the dimensionality of one's solution. One might check the stability of solutions of different dimensionalities and choose one's final solution only from those solutions that are sufficiently stable.

Apart from testing INDOMIX on several data sets, in passing also the varimax procedure for PCAMIX has been used consistently. The usefulness of this procedure has been discussed in one case only. It has been used in every analysis, however, and seemed very useful. Especially when one wants to interpret the components, it is useful to have a simple structure for the loadings.

It may seem rather inconsistent to use the varimax criterion for rotating the MCA solution and compare this solution with INDOMIX, which

maximizes the quartimax criterion. The reason for this apparent inconsistency is that the varimax criterion is preferred over the quartimax criterion, for the arguments given by Kaiser (1958), but the variant of INDOMIX that maximizes the varimax criterion has not yet been programmed. Moreover, INDOMIX itself has an interesting interpretation as a compromise between MCA and PCA of quantification matrices. This interpretation does not hold for the varimax based variant of INDOMIX.

The final example has not only been taken up to show the limitations of INDOQUAL (and in fact also of MCA). It also shows that analyzing one's data by just one method may hide important aspects of the data. A more useful strategy seems to be to use more than one of the methods mentioned in the hierarchy, possibly even all of these. Then one should not only consider the solutions of each of the separate analyses, but especially the larger differences in the solutions. Together with the knowledge how the methods presented in this study are related, one may determine whether or not the data can be described by these methods, and if so, which is the most useful representation of this data.