CHAPTER 5

EXPERIMENT 1, FIRST COMPARISON OF THE THREE SCA-METHODS

5.1. INTRODUCTION

In the first experiment, conducted in this study, the objective was to differentiate between the methods SCA–W, SCA–P and SCA–S. For this, the performance of the three methods on the various measures and success criteria under a variety of circumstances was investigated. For this reason, the independent variables were manipulated over wide ranges of values, although some common characteristics were maintained throughout all conditions, and values of some independent variables were kept within realistic bounds.

5.2 MANIPULATIONS OF THE DATA

The data (correlation matrices), analyzed in this experiment, were constructed as described in Section 3.2. The following values were chosen for the independent variables. In all data sets there were two samples (p=2), 12 or 24 variables (=m), 2 or 4 factors (=q) and sample sizes of n=50, n=100 or n=150. The true pattern matrices ($\mathbf{P}_a$ to $\mathbf{P}_g$) used in Experiment 1 are given in Appendix A. These are used in twelve different combinations (see Sections 5.2.1 and 5.2.2), six of which lead to one–population data and six to two–or–more–populations data. In total six different true correlation matrices $\mathbf{\Phi}_t$, with correlations between the factors were used. The factors were defined to be orthogonal ($\mathbf{\Phi}_t$=$\mathbf{I}$), mildly oblique ($\mathbf{\Phi}_t$=$\mathbf{\Phi}_{.2}$) or strongly oblique ($\mathbf{\Phi}_t$=$\mathbf{\Phi}_{.4}$; see Section 4.1.2).

Now the rationale for the choice of the specific true pattern matrices used will be given. The first three true pattern matrices $\mathbf{P}_a$, $\mathbf{P}_b$ and $\mathbf{P}_c$ define strong factors, intermediate factors and weak factors, respectively. With these matrices the overall effect of factor strength on the dependent variables was investigated. The true pattern matrices $\mathbf{P}_d$ and $\mathbf{P}_e$ define factors of moderately different and strongly different strengths, respectively. With these matrices it was investigated whether the retrieval of weak factors was influenced by the strengths of the other factors in the population. The true pattern matrices $\mathbf{P}_f$ and $\mathbf{P}_g$, finally, define factors of equal strengths, with unequal loadings per factor, descending and ascending in strength, respectively. These matrices were used to simulate two-or-more-populations data, described in Section 3.3. The combinations of these seven different true pattern matrices used form a reasonable representation of the different situations one might expect to encounter in the "real life" research domain.

In each condition, described by choices of the number of variables in the population (NVar), the number of factors in the population (NFac), sample size (SampSize), combinations of pattern matrices (PatternComb) and combinations of correlation matrices $\mathbf{\Phi}_t$ (PhiComb), ten replications of simulated data were used. The methods of analysis used were SCA-W, SCA-P, SCA-S and (only for comparison) PCA-sep. Now the pattern combinations leading to one-population data and two-or-more-populations data will be described.

5.2.1 One-population data

In one-population data, the columns of the true pattern matrices of the two groups are perfectly congruent. The same goes for the columns of the two true structure matrices. So the CM across groups for all true pattern matrices and all true structure matrices is 1.00 (see Section 3.3.1).

The combinations of true pattern matrices $\mathbf{P}_k$ (see Appendix A) and matrices $\mathbf{\Phi}_k$, used to simulate the two groups in each data set for one-population data, were '$\mathbf{P}_a\mathbf{P}_a$', '$\mathbf{P}_b\mathbf{P}_b$', '$\mathbf{P}_c\mathbf{P}_c$', '$\mathbf{P}_d\mathbf{P}_d$', '$\mathbf{P}_e\mathbf{P}_e$', '$\mathbf{P}_a\mathbf{P}_c$' and '$\mathbf{I}$ $\mathbf{I}$', '$\mathbf{\Phi}_{.2}\mathbf{\Phi}_{.2}$' and '$\mathbf{\Phi}_{.4}\mathbf{\Phi}_{.4}$', respectively. It can be verified that all combinations indeed satisfy the definition of one-population data mentioned above. Within each combination of true pattern matrices and correlation matrices (18 combinations in total), data sets were simulated using 12 or 24 variables, 2 or 4 factors and sample sizes of 50, 100 and 150, amounting to 216 different conditions within the category of one–population data to be analyzed.

5.2.2 Two–or–more–populations data

In two–or–more–populations data, the true pattern and the true structure matrices are *not* proportional over the groups. The scores for two simulated groups are generated differently. That is, for each group a different true pattern matrix is used.

The combinations of true pattern matrices $\mathbf{P}_k$ (see Appendix A) and matrices $\mathbf{\Phi}_k$, used to simulate data for the two groups in each data set for two–or–more–populations data are given in Table 5.1, together with the value of the CM between the two structure matrices for each combination. The CM is reported to display the similarity between the two true structure matrices in each condition. The CM for the true pattern matrices is independent of the choice of matrices $\mathbf{\Phi}_k$, so that the values of the CM are the same for all choices of matrix $\mathbf{\Phi}_k$. They can be found on the first row of Table 5.1. From Table 5.1 it can be seen that, when the correlation matrices $\mathbf{\Phi}_k$ are chosen differently for each group, the CM for the true structure matrices drops considerably below the CM for the true pattern matrices.

Within each combination of true pattern matrices and correlation matrices (30 combinations in total), data sets were simulated using 12 or 24 variables, 2 or 4 factors and sample sizes of 50, 100 and 150, amounting to 360 different conditions within the category of two–or–more–populations data to be analyzed.

5.3 DEPENDENT VARIABLES

For each data set of two groups, the amount (percentage) of variance explained was calculated for 1, 2, 3, 4, 5 and 6 components drawn. In the

Table 5.1: *The CM for the true structure matrices*[1] *for combinations of true pattern matrices* $\mathbf{P}_k$ *and matrices* $\mathbf{\Phi}_k$, *used for two-or-more-populations data*

q	PhiComb condition	$\mathbf{P}_a\mathbf{P}_g$	$\mathbf{P}_b\mathbf{P}_g$	$\mathbf{P}_c\mathbf{P}_g$	$\mathbf{P}_d\mathbf{P}_g$	$\mathbf{P}_e\mathbf{P}_g$	$\mathbf{P}_f\mathbf{P}_g$
2	$\mathbf{I}\ \mathbf{I}$	.94	.94	.94	.94	.94	.78
2	$\mathbf{\Phi}_{.2}\mathbf{\Phi}_{.2}$	.94	.94	.94	.94	.92	.78
2	$\mathbf{\Phi}_{.4}\mathbf{\Phi}_{.4}$	.94	.94	.94	.93	.90	.78
2	$\mathbf{I}\ \mathbf{\Phi}_{.4}$	.87	.87	.87	.87	.87	.72
2	$\mathbf{\Phi}_{.4}\ \mathbf{I}$	.87	.87	.87	.85	.80	.72
4	$\mathbf{I}\ \mathbf{I}$	.94	.94	.94	.94	.94	.78
4	$\mathbf{\Phi}_{.2}\mathbf{\Phi}_{.2}$	.94	.94	.94	.93	.91	.78
4	$\mathbf{\Phi}_{.4}\mathbf{\Phi}_{.4}$	.94	.94	.94	.92	.88	.78
4	$\mathbf{I}\ \mathbf{\Phi}_{.4}$	.77	.77	.77	.77	.77	.64
4	$\mathbf{\Phi}_{.4}\ \mathbf{I}$	.77	.77	.77	.75	.70	.64

[1]*The CM for the true pattern matrices is the same for all choices of matrix* $\mathbf{\Phi}_k$ *and is given on the first row* (**I I**) *for* $q=2$ *and* $q=4$.

4 factor condition, the explained variance for 7 components drawn was also calculated. This was done for the measures QA and QDA. In order to get the QDA value for q components, the explained variance of $q+2$ components (among other things) is required (see Section 4.2.4). If the maximum number of components for which the explained variance is calculated in the 4 factors condition would be 6, the maximum number of components that could be indicated by the measure QDA would be 4, so no overestimation could occur. For this reason, the amount of variance explained by 7 components was also calculated in the $q=4$ condition. The 7 components solution was used for the measure QA (making possible an indication of 6 components) and QDA (making possible an indication of 5 components). For the dimension indicators KA1 and PA, the 7 component solution was not of interest, because, for the success rate of the dimension indicators it was only inspected whether or not overindication occurred, regardless of how large the overindication was.

According to the description of the dimension indicators, given in Section 4.2.5, it was recorded for each data set whether under–, correct or overestimation of the number of factors occurred. For comparison, the results of the dimension indicators for PCA-sep were also included.

For each solution with the correct dimension, the measures RR and DFC were calculated (see Section 4.4.1. and Section 4.4.2). As far as the measure RR is concerned, the main focus was on the RR for the pattern matrix for SCA-P and SCA-W, and the RR for the structure matrix for SCA-S.

The simulated data used in the present experiment could be divided into two main categories: Data coming from one population (one-population data) and data coming from two (different) populations (two-or-more-populations data). In the present experiment, we investigated whether the data from these two categories could be distinguished by using a discriminant analysis with the variables SampSize, NVar, NFac and the percentages of explained variance of the SCA-methods and PCA-sep as predictors. The number of misclassifications will be reported, together with the discriminant function.

5.4 ANALYSIS

All the results presented were analyzed with repeated measures analyses of variance. Sums of squares of main and interaction effects were calculated to give an impression of the size of the effects, and the associated averaged univariate F tests were employed for significance testing. Multivariate criteria such as Wilks' lambda were also checked and it was found that using this criterion led to similar results. So employing these other multivariate criteria would not change the conclusions presented here.

For the analyses, it had to be checked whether or not assumptions were violated. The three assumptions for the F tests in a univariate repeated measures analysis are: Independence of observations, multivariate normality and sphericity. The procedure for simulating the data, used in the present study, guarantees that the assumption of independence of observations is fulfilled. Furthermore, the analysis is fairly robust against violation of the assumption of normality (for the present study it is obvious that this assumption is violated) and, because the α used is chosen to be very small (α=.001), little harm can be expected from such violations. Finally, to correct for possible nonsphericity, the adjustment of the number of degrees of freedom, proposed by Greenhouse and Geisser (1959), was used.

The analyses were carried out using the MANOVA module in SPSS-PC 5.01 extended. The analyses were performed on the full data set, and on the data sets for one-population data and two-or-more-populations data, separately. The between subjects variables in the analyses of the complete data set were: Population (2 conditions, one-population data and two-or-more-populations data – note that for each population condition there were 6 different conditions of PatternComb), PhiComb (3 or 5 conditions), SampSize (3 conditions; n=50, 100, 150), NVar (2 conditions; 12 and 24 variables), and NFac (2 conditions; 2 and 4 factors). In the separate analyses of the samples from one population (one-population data) and the samples from two populations (two-or-more-populations data), the between subjects variable PatternComb was included in the analysis. In all analyses, the variable Method was specified as a within subjects factor (3 conditions; SCA-W, SCA-P and SCA-S) in a repeated measures design, because each method of analysis was presented with the same data sets to be analyzed. In each of the separate analyses, only first and second order interactions among between factors were included in the analysis. Prior analyses with a full factorial model showed that there were no relevant higher order interactions present.

For the analysis of the results for the dimension indicators, PCA-sep was included as a condition of the within subjects factor, so in that case the within subjects factor had 4 conditions. In order to be able to perform a MANOVA, observations concerning correct dimension indication were aggregated over the ten replications within each condition to obtain a reasonably continuous variable: The success rate of indicating the correct number of components over ten replications. This means that for each dimension indicator for each method, for each condition a value between 0 and 1.0 was obtained. In a first analysis, the dimension indicator used was specified as a within subjects factor (4 conditions; KA1, PA, QA and QDA) in a repeated measures design. After choosing the preferred dimension indicator, only the results for this measure were used.

5.5 QUESTIONS TO BE ANSWERED BY THIS EXPERIMENT

The present experiment was conducted to get answers as to what factors play a role in the recovery of factors by the methods at hand, and which measures can best be used for determining the number of components present in a data set. These questions are specified as below. Answers will be given in separate subsections of Section 5.6, where all results will be presented. The question numbers refer to the subsections in which answers to these questions will be given. Summary conclusions from the results of this experiment will be drawn in Section 5.7.

1.) Which of the dimension indicators KA1, PA, QA and QDA can best be used by each of the SCA-methods to indicate the number of components to retain?

2a.) Which of the SCA-methods has the highest success rate on the preferable dimension indicator?

2b.) Are there interesting interaction effects between independent variables and method of analysis for the most preferable dimension indicator?

3a.) Which of the SCA-methods can be judged as best capable of recovering factors?

3b.) Are there interesting interaction effects between independent variables and method of analysis for the recovery of factors?

4a.) Which of the SCA-methods can be judged as best capable of recovering the correlations between factors?

4b.) Are there interesting interaction effects between independent variables and method of analysis for the recovery of correlations between factors?

5.) Can samples from one population be distinguished from samples from two populations?

Besides the above questions, a few questions concerning the effect of specific choices of the pattern matrices were asked.

6a.) Which SCA-method gives the highest RR, when comparing samples taken from one population, with only strong factors (condition '$\mathbf{P}_a\mathbf{P}_a$'), only intermediate factors (condition '$\mathbf{P}_b\mathbf{P}_b$'), and only weak factors (condition '$\mathbf{P}_c\mathbf{P}_c$')?

6b.) Are, for the SCA-methods, intermediate factors and weak factors blurred by strong factors, or vice versa? For this purpose, the conditions '$\mathbf{P}_d\mathbf{P}_d$' and '$\mathbf{P}_e\mathbf{P}_e$' were compared with conditions '$\mathbf{P}_b\mathbf{P}_b$' and '$\mathbf{P}_c\mathbf{P}_c$', respectively.

6c.) Is the RR for the three SCA-methods affected by the existence of small measurement error in one sample and large measurement error in another sample (condition '$\mathbf{P}_a\mathbf{P}_c$'), and which SCA-method performs best under this condition? For this purpose, results for condition '$\mathbf{P}_a\mathbf{P}_c$' were compared with those for condition '$\mathbf{P}_a\mathbf{P}_a$' and condition '$\mathbf{P}_c\mathbf{P}_c$'.

5.6 RESULTS

In the present experiment, comparisons were made with large numbers of observations in each condition, which, in many cases, caused even small differences to become significant. Because small differences have little practical value, these will be ignored. Overall, a significance level α=.001 was used. Whenever differences between dependent variables or methods are reported, or when interaction effects are said to occur, effects were significant at α=.001.

For each of the conditions of independent variables, used in the figures, the number of data sets analyzed on which the reported values of the success criteria are based, is given in Table 5.2. In the following Sections, each of the six main questions, asked in of Section 5.5, will be answered separately. At the start of each Section, the question(s) will be repeated.

Table 5.2: *The number of observations for the conditions of the independent variables on which the values in the figures are based.*

Independent Variable	all samples	n	1 population	n	2 populations	n
All		5760				
Population	1 pop	2160	1 pop	2160	2 pop	3600
	2 pop	3600				
PatternComb	See breakdown over 1 and 2 populations		$\mathbf{P}_a\mathbf{P}_a$	360	$\mathbf{P}_a\mathbf{P}_g$	600
			$\mathbf{P}_b\mathbf{P}_b$	360	$\mathbf{P}_b\mathbf{P}_g$	600
			$\mathbf{P}_c\mathbf{P}_c$	360	$\mathbf{P}_c\mathbf{P}_g$	600
			$\mathbf{P}_d\mathbf{P}_d$	360	$\mathbf{P}_d\mathbf{P}_g$	600
			$\mathbf{P}_e\mathbf{P}_e$	360	$\mathbf{P}_e\mathbf{P}_g$	600
			$\mathbf{P}_a\mathbf{P}_c$	360	$\mathbf{P}_f\mathbf{P}_g$	600
PhiComb	$\mathbf{I}\ \mathbf{I}$	1440	$\mathbf{I}\ \mathbf{I}$	720	$\mathbf{I}\ \mathbf{I}$	720
	$\mathbf{\Phi}_{.2}\mathbf{\Phi}_{.2}$	1440	$\mathbf{\Phi}_{.2}\mathbf{\Phi}_{.2}$	720	$\mathbf{\Phi}_{.2}\mathbf{\Phi}_{.2}$	720
	$\mathbf{\Phi}_{.4}\mathbf{\Phi}_{.4}$	1440	$\mathbf{\Phi}_{.4}\mathbf{\Phi}_{.4}$	720	$\mathbf{\Phi}_{.4}\mathbf{\Phi}_{.4}$	720
	$\mathbf{I}\ \mathbf{\Phi}_{.4}$	720			$\mathbf{I}\ \mathbf{\Phi}_{.4}$	720
	$\mathbf{\Phi}_{.4}\ \mathbf{I}$	720			$\mathbf{\Phi}_{.4}\ \mathbf{I}$	720
SampSize	$n=50$	1920	$n=50$	720	$n=50$	1200
	$n=100$	1920	$n=100$	720	$n=100$	1200
	$n=150$	1920	$n=150$	720	$n=150$	1200
NVar	$m=12$	2880	$m=12$	1080	$m=12$	1800
	$m=24$	2880	$m=24$	1080	$m=24$	1800
NFac	$q=2$	2880	$q=2$	1080	$q=2$	1800
	$q=4$	2880	$q=4$	1080	$q=4$	1800

5.6.1 Finding the preferable dimension indicator

Question 1: "Which of the dimension indicators KA1, PA, QA and QDA can best be used by each of the SCA-methods to indicate the number of components to retain?"

For this, the four dimension indicators were compared for each of the methods separately. In Figure 5.1a to 5.1d, the percentage of times the correct number of components was indicated by the four dimension indicators is presented for each method separately. The percentages are given, summed over all conditions taken together, and summed over each condition of each independent variable separately.

For PCA–sep, the measure QDA gave significantly higher percentages of correct dimension indications (69.4%) than the measures QA (57.4%), PA (58.4%), and KA1 (35.5%). As can be seen from Figure 5.1a, the measure QDA also has the highest percentage of correct dimension indications of the four dimension indicators, when summed over each condition of each independent variable separately.

For SCA–W, the same holds (see Figure 5.1b). The measure QDA gave significantly higher percentages of correct dimension indications (74.6%) than the measures QA (66.3%), PA (61.6%) and KA1 (45.6%), and has the highest percentage of correct dimension indications of the four dimension indicators, when summed over each condition of each independent variable separately.

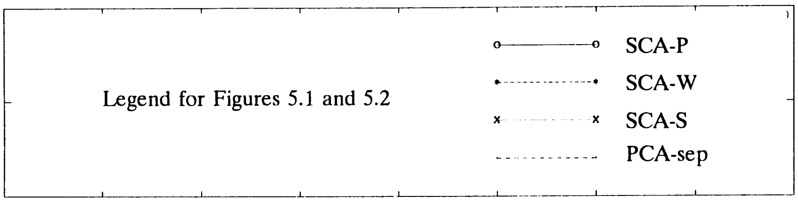
For SCA–P, the measure QDA also gave significantly higher percentages of correct dimension indications (75.3%) than the measures PA (61.8%), and KA1 (51.1%). The difference with the measure QA (69.7%) is not significant. As can be seen from Figure 5.1c, the measure QDA does not always have the highest percentage of correct dimension indications of the four dimension indicators, when summed over each condition of each independent variable separately. For instance, when 12 variables are used, the measure KA1 has a higher percentage of correct dimension indications, although differences are small.

Finally, for SCA–S, the story is somewhat different. For this method, the measure QA gives the highest percentages of correct dimension indications (55.8%), but the differences with the measure QDA (51.0%) is not significant in contrast to the differences with the measures PA (43.8%) and KA1 (38.1%). The measure QDA also has significantly higher percentages of correct dimension indications than the measures PA and KA1.

In Figure 5.2, the percentage of times the correct number of components was indicated by each dimension indicator and each method, together with the percentages of times too many or too few components were indicated, are presented. The measure KA1 almost never indicated too few components, but it did indicate too many components more often than the correct number of components for all methods except SCA–P.

The measure PA indicated the correct number of components more often than it over– or underindicated the number of components. Both over– and underindication occurred.

The measures QA and QDA indicated the correct number of components more often than they over– or underindicated the number of components. Both measures underindicated the number of components, but overindication did not occur much, except for SCA–S.



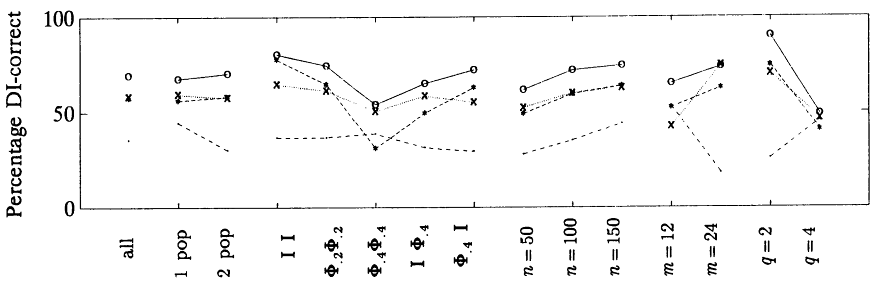


Figure 5.1a *Percentages of correct dimension indications for the method PCA – sep, averaged over each level of each independent variable, using four different dimension indicators*

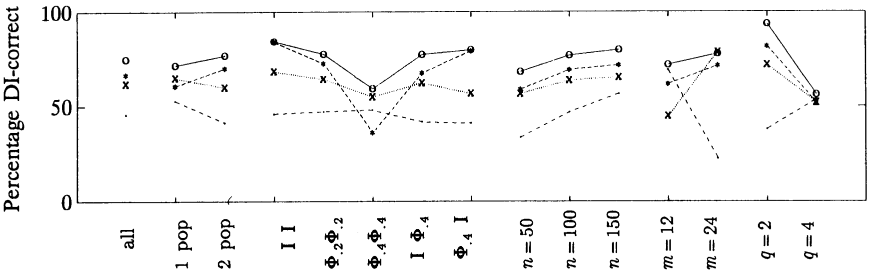


Figure 5.1b *Percentages of correct dimension indications for the method SCA-W, averaged over each level of each independent variable, using four different dimension indicators*

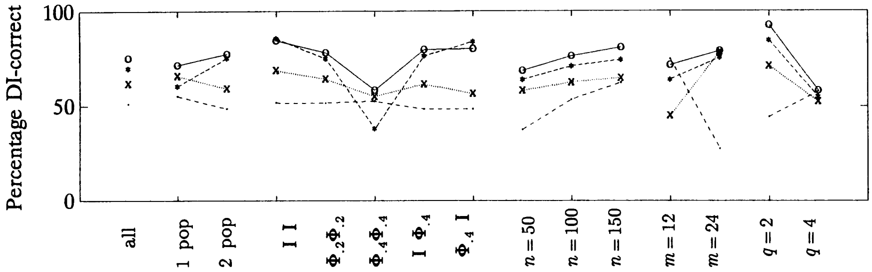


Figure 5.1c *Percentages of correct dimension indications for the method SCA-P, averaged over each level of each independent variable, using four different dimension indicators*

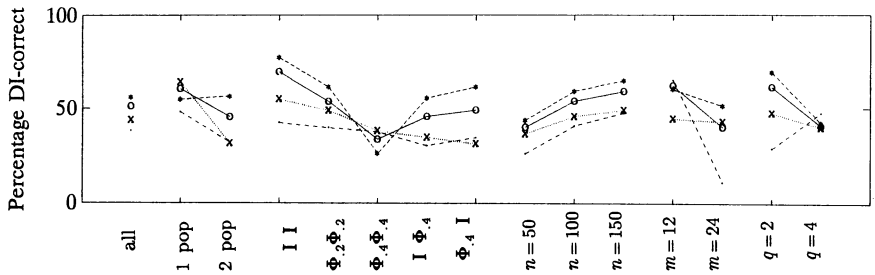


Figure 5.1d *Percentages of correct dimension indications for the method SCA-S, averaged over each level of each independent variable, using four different dimension indicators*

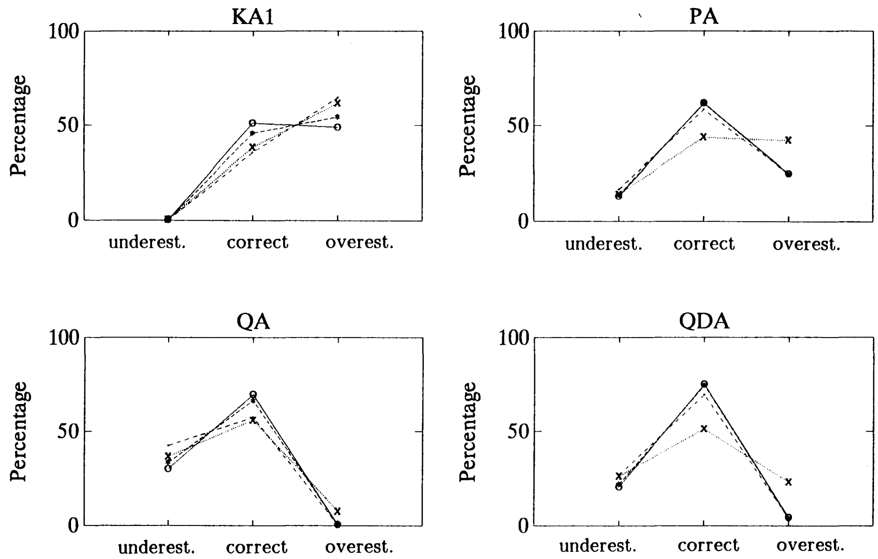


Figure 5.2 *Percentages of times the correct number of components was indicated by each method for four dimension indicators, and the percentages of times over and underestimation occurred*

Answer 1: Because of the superiority of the measure QDA in indicating the correct number of components summed over all conditions of the independent variables and over all methods of analysis, for the remainder of this study the measure QDA was used for dimension indication.

5.6.2 Finding the SCA-method with the best dimension indication

Question 2a: "Which of the SCA-methods has the highest success rate on the preferable dimension indicator?"

Question 2b: "Are there interesting interaction effects between independent variables and method of analysis for the most preferable dimension indicator?"

To get an overview of the size of the main effect of Method and its interactions with one or two between subject factors on the success of the dimension indicator QDA, the associated sums of squares for the successrate of QDA in the conditions 'samples from one population' and 'samples from two populations', are presented in Table 5.3, where asterisks indicate effects significant at α=.001.

Overall, the four methods differed significantly on the QDA success rate. SCA–P had a significantly higher success rate for the measure QDA (75.3%) than SCA–S (51.0%)and PCA–sep (69.4%), and was also better than SCA–W (74.6%), albeit not significantly. There was a significant interaction effect of Population × Method: The measure QDA had lower success rates in the 'one population' condition than in the 'two populations' condition for SCA–P (71.7% vs. 77.5%), SCA–W (71.4% vs. 76.6%) and PCA–sep (67.8% vs. 70.4%), while for SCA–S the opposite was

Table 5.3: *Sums of Squares for the measure QDA from the ANOVA analysis of samples from one population* ($n = 2160$) *and samples from two populations* ($n = 3600$)

	one population	two populations
Total within subjects	12.00	67.19
Method	1.76*	24.46*
PatternComb × Method	1.40*	6.18*
PhiComb × Method	.10	1.22*
SampSize × Method	.52*	.12
NVar × Method	.67*	10.59*
NFac × Method	.03	5.90*
PatternComb × PhiComb × Method	.65*	1.99*
PatternComb × SampSize × Method	.67*	.32
PatternComb × NVar × Method	.61*	1.54*
PatternComb × NFac × Method	.98*	2.40*
PhiComb × SampSize × Method	.06	.32
PhiComb × NVar × Method	.09	.09
PhiComb × NFac × Method	.11	.72*
SampSize × NVar × Method	.11	.26
SampSize × NFac × Method	.03	.08
NVar × NFac × Method	.05	.72*
Residual within subjects	4.15	10.28

Method = Method used, PatternComb = Used Pattern Combinations, PhiComb = Used Combinations of matrices Φ*, NFac = Number of factors, NVar = Number of variables. * = significant at* α *= .001.*

the case (60.5% vs. 45.4%). Because of this population effect, and because of the incomparability of the patterns used in those two conditions, the results for the QDA measure in both population conditions were inspected.

In Figure 5.3a and 5.3b, the percentages of correct dimension indications by the measure QDA are presented for the SCA-methods and PCA-sep, summed over each condition of each independent variable, for samples from one population and samples from two populations, respectively.

From Figure 5.3a, it can be seen that SCA-W and SCA-P gave similar results for all conditions of the independent variables, while the difference in success of the dimension indicator QDA for SCA-S and PCA-sep compared to SCA-W and SCA-P was relatively small for some conditions of the independent variables and large for other conditions of the independent variables. For the samples from one population, among the independent variables, the greatest differences in the overall success of the dimension indicator QDA are caused by the variable PatternComb. For the conditions '$\mathbf{P}_c\mathbf{P}_c$' and '$\mathbf{P}_e\mathbf{P}_e$' (these are the conditions where the largest amounts of error were present in the data), the dimension indicator QDA performs very poorly, compared to the other conditions of PatternComb. The interaction of PatternComb × Method was the strongest interaction effect, although it explained only 11.7% of the variation (see Table 5.3: Sums of squares of PatternComb × Method is 1.40, see also Figure 5.3a). The effect with the highest sums of squares was Method, explaining 14.7% of the total within subjects variation. The methods SCA-P and SCA-W had the highest percentages of correct dimension indications, SCA-S the lowest, and PCA-sep scored in between. All other sums of squares were considered too small to be relevant.

The results for the samples from two populations (Figure 5.3b) give about the same picture. The success of the dimension indicator for SCA-W and for SCA-P is about the same. This was the case for all conditions of the independent variables, while the difference in success of the dimension indicator QDA for SCA-S and PCA-sep compared to SCA-W and SCA-P was relatively small for some conditions of the independent variables and large for other conditions of the independent variables. From Table 5.3, we see that the four most relevant effects are the effect of Method and the interactions of Method × PatternComb, Method × NVar, and Method × NFac. The effect of Method is the strongest effect, just as for the samples from one population, explaining 36.4% of the total variation. The interaction of Method × PatternComb can be found, for instance, in small differences between SCA–S and the other methods in PatternComb condition '$\mathbf{P}_c\mathbf{P}_g$' and large differences in '$\mathbf{P}_f\mathbf{P}_g$'. The interaction of Method × NVar comes from the fact that SCA–S performs worse when 24 variables are present than when 12 variables are present, while SCA–W and SCA–P perform better for 24 variables than for 12 variables. The interaction of Method × NFac comes from the fact that the success rate of the measure QDA is about 42% and 36% lower for SCA–W and SCA–P, respectively, when 4 factors are present than when 2 factors are present, while this difference is only about 14% for SCA–S. Together these four effects explain 70.1% of the within subjects variation.

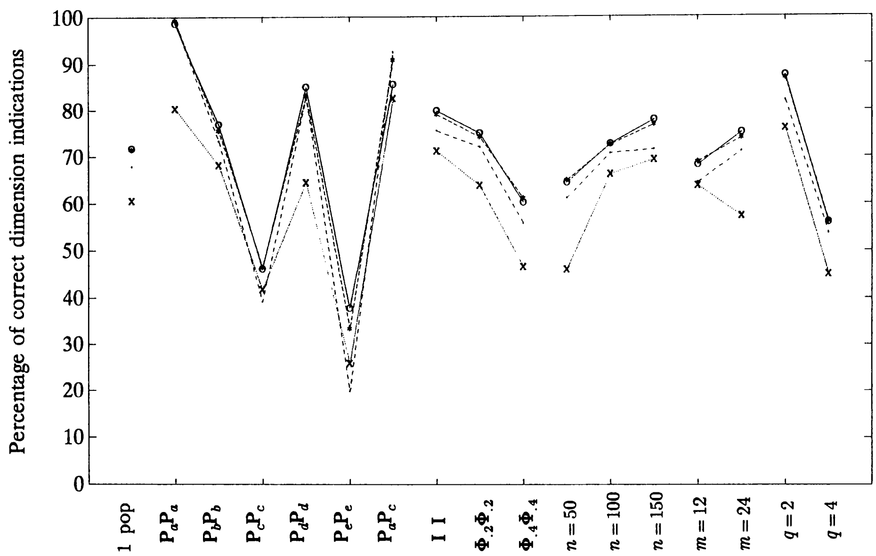


Figure 5.3a *Percentages of correct dimension indications for the three SCA–methods and PCA–sep, averaged over each level of each independent variable, within the samples from one population, using the dimension indicator QDA*

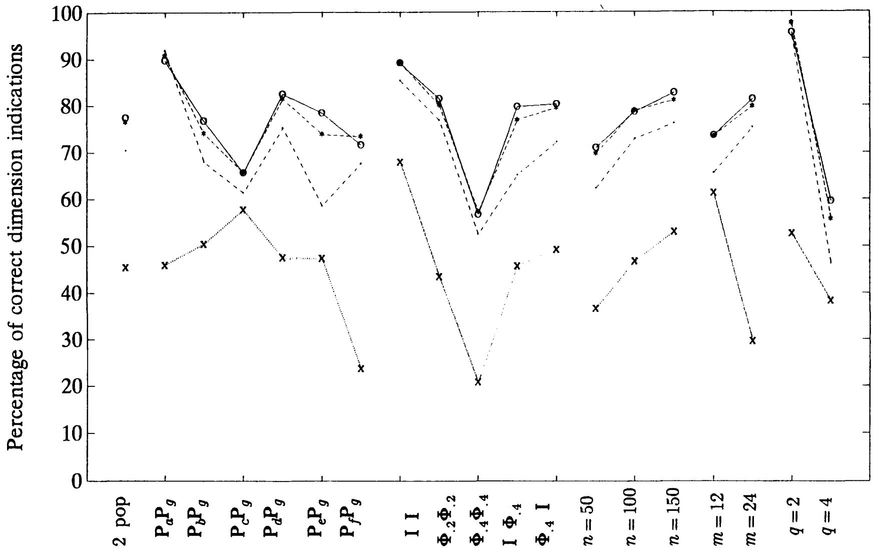
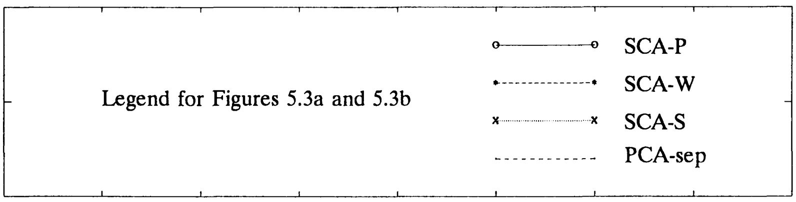


Figure 5.3b *Percentages of correct dimension indications for the three SCA–methods and PCA–sep, averaged over each level of each independent variable, within the samples from two populations, using the dimension indicator QDA*



Concluding, we can say that both in 'samples from one population' and in 'samples from two populations', there was an interesting interaction of PatternComb × Method. The measure QDA indicated the correct number of components much less often for SCA–S than for SCA–W and SCA–P in some of the PatternComb conditions from 'samples from one population' (especially '$\mathbf{P}_a\mathbf{P}_a$' and '$\mathbf{P}_d\mathbf{P}_d$') and in some of the PatternComb conditions from 'samples from two populations' (especially '$\mathbf{P}_a\mathbf{P}_g$' and '$\mathbf{P}_f\mathbf{P}_g$'), while in some of the other PatternComb conditions, success rates for the three SCA–methods hardly differed.

The interaction effect of Population × Method, noted at the beginning of this Section, can now be seen to be mainly caused by the choices of the different conditions of PatternComb. The two conditions of PatternComb with large amounts of error, used in the condition 'samples from one population', caused the relatively poor performance for SCA–W, SCA–P and PCA–sep, thus giving the impression that the number of components were better indicated for samples from two populations than for samples from one population. However, because the conditions of PatternComb in samples from one population do not have a clear counterpart in the samples from two populations, this conclusion can not be assumed to hold generally. For SCA–S, it can be said that it gave *relatively* better results for the samples from one population than for the samples from two populations, indicating that beside the amount of error, the CM of the true structure matrices is also an important factor in determining the dimensionality correctly.

Answer 2a: From the results, presented above, it was concluded that SCA–W and SCA–P were most successful in indicating the correct number of components, using the measure QDA. These two methods are therefore recommended, when the correct number of components has to be determined solely from the data, that is, when the number of components can not be decided on on theoretical grounds. SCA–P has our special preference, because it takes much less time to execute.

Answer 2b: There was an interesting interaction of PatternComb × Method, caused by the different behavior of the measure QDA for SCA–S, compared to the other two SCA-methods. For data from two populations, there was also an interesting interaction of Method × NVar, again caused by SCA–S. SCA–S gives poorer results when 24 variables are present than when 12 variables are present, while for SCA–W and SCA–P, the opposite is the case.

5.6.3 Finding the SCA-method with the best recovery of factors

Question 3a: "Which of the SCA–methods can be judged as best capable of recovering factors?"

Question 3b: "Are there interesting interaction effects between independent variables and method of analysis for the recovery of factors?"

In the present experiment, the recovery rate (RR) was computed both using the pattern matrices and using the structure matrices. Whereas the RR was smaller in samples from one population than in samples from two populations for SCA–W and SCA–P (for both the pattern and the structure matrix), the opposite was the case for SCA–S (for both the pattern and the structure matrix). When looking at the RR for the pattern matrices, found with SCA–W and SCA–P, and the structure matrix, found with SCA–S, there was a significant interaction effect of Population × Method on the RR. Therefore, results for the conditions 'samples from one population' and 'samples from two populations' will again be presented separately.

Firstly, the RR's of the pattern matrix and the structure matrix will be presented for each method separately, to show that the use of the pattern matrix for SCA–W and SCA–P and the structure matrix for SCA–S in the final comparison can be justified on empirical grounds, that is, the matrices used had the best overall performance. Secondly, the methods will be compared using the RR measure.

In Figures 5.4a, 5.4b and 5.4c, the RR's of the pattern and the structure matrix are presented for each method separately, summed over each condition of the independent variables, within the one–population data. In Figure 5.5a to 5.5c, the same is done for two–populations data.

For SCA–W and SCA–P, the differences between the RR of the pattern and the structure matrix are small, the pattern matrix being the matrix with the highest overall RR's. For SCA–S, the RR of the structure matrix is the highest for all conditions of the independent variables, within the condition 'samples from one population'. Within the condition 'samples from two populations', the overall RR of the structure matrix is still higher than the RR of the pattern matrix, but for some conditions of the independent variables, the pattern matrix has a higher RR. This is not very surprising, because, when the samples come from different populations, there is not one single structure matrix to be found.

On the basis of these results, it is decided that in the remainder of

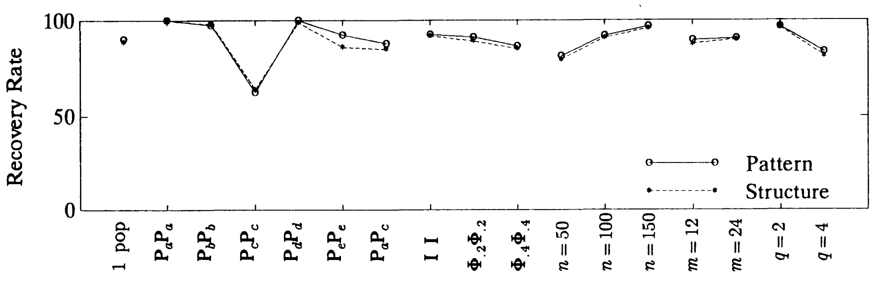


Figure 5.4a *Recovery Rates for the method SCA−W, based on the pattern and the structure matrix, averaged over each level of each independent variable, within the samples from one population*

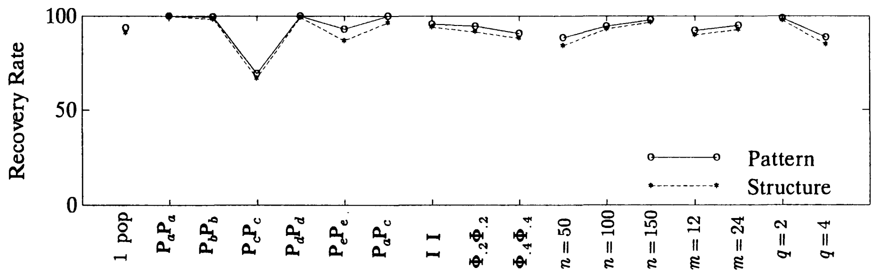


Figure 5.4b *Recovery Rates for the method SCA−P, based on the pattern and the structure matrix, averaged over each level of each independent variable, within the samples from one population*

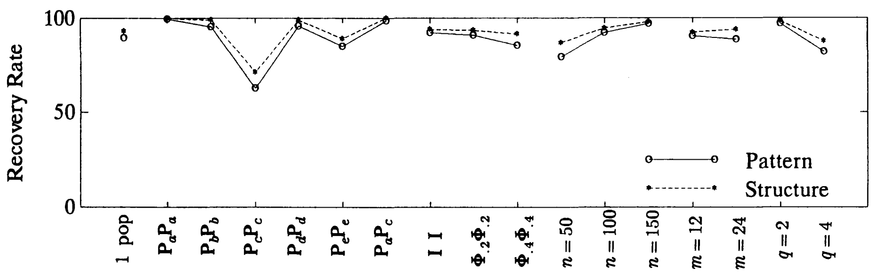


Figure 5.4c *Recovery Rates for the method SCA−S, based on the pattern and the structure matrix, averaged over each level of each independent variable, within the samples from one population*

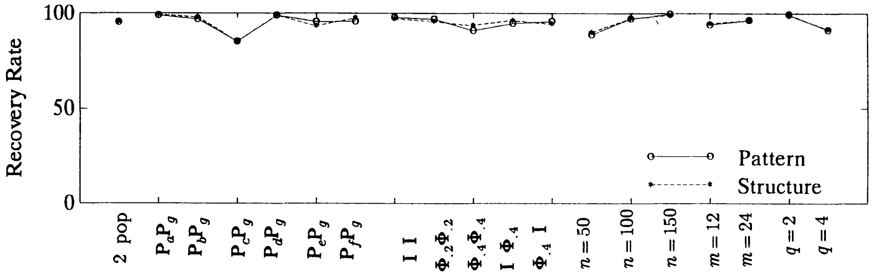


Figure 5.5a *Recovery Rates for the method SCA-W, based on the pattern and the structure matrix, averaged over each level of each independent variable, within the samples from two populations*

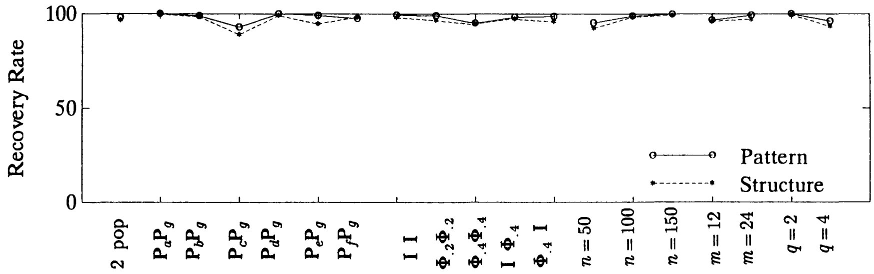


Figure 5.5b *Recovery Rates for the method SCA-P, based on the pattern and the structure matrix, averaged over each level of each independent variable, within the samples from two populations*

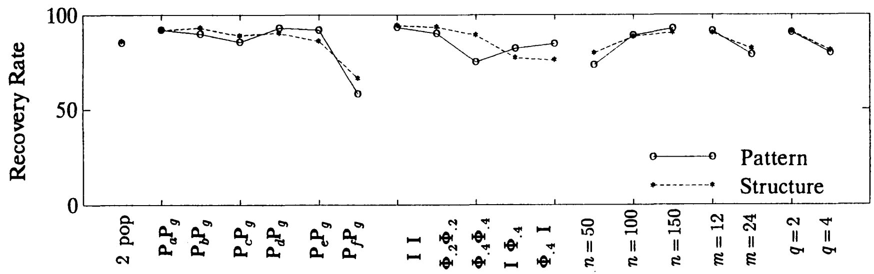


Figure 5.5c *Recovery Rates for the method SCA-S, based on the pattern and the structure matrix, averaged over each level of each independent variable, within the samples from two populations*

this study the RR of the pattern matrix, found with SCA–W, will be compared with the RR of the pattern matrix, found with SCA–P and the RR of the structure matrix, found with SCA–S, as index of factor recovery.

To get an overview of the size of the main effect of Method and its interactions with one or two between subject factors, the associated sums of squares for the success criterion RR in the conditions 'samples from one population' and 'samples from two populations', are presented in Table 5.4, where significant effects at α=.001 are marked by asterisks. The most important effects will be described in Sections 5.6.3.1 and 5.6.3.2.

Table 5.4: *Sums of Squares for the measure RR from the ANOVA analysis of samples from one population* ($n = 2160$) *and samples from two populations* ($n = 3600$)

	one population	two populations
Total within subjects	36.23	147.53
Method	1.67*	27.47*
PatternComb × Method	3.81*	23.03*
PhiComb × Method	.24*	15.32*
SampSize × Method	.86*	1.74*
NVar × Method	.07	6.80*
NFac × Method	.33*	1.76*
PatternComb × PhiComb × Method	.47*	3.06*
PatternComb × SampSize × Method	2.02*	1.59*
PatternComb × NVar × Method	.15	10.40*
PatternComb × NFac × Method	.92*	2.70*
PhiComb × SampSize × Method	.15	1.09*
PhiComb × NVar × Method	.07	.81*
PhiComb × NFac × Method	.07	5.56*
SampSize × NVar × Method	.09	.39*
SampSize × NFac × Method	.07	.62*
NVar × NFac × Method	.11*	.04
Residual within subjects	25.13	45.18

Method = Method used, PatternComb = Used Pattern Combinations, PhiComb = Used Combinations of matrices Φ, *NFac = Number of factors, NVar = Number of variables. * = significant at* α = .001.

5.6.3.1 *Samples from one population*

In Figure 5.6, the RR's are presented, summed over each condition of each independent variable, for the pattern matrix, found with SCA–W and SCA–P, and for the structure matrix, found with SCA–S. Overall, a significant Method effect existed (see Table 5.4). The methods SCA–P and SCA–S both had a significantly higher RR (93.6% and 93.1%, respectively) than SCA–W (90.0%). The difference between SCA–P and SCA–S was not significant at α=.001 (p=.015). Summed over each condition of the independent variables SampSize, NVar and NFac separately (these are the independent variables known in practice), SCA–W always had the lowest RR.

The largest effect (see Table 5.4) was the interaction of PatternComb × Method. Looking at the six conditions of the independent variable PatternComb, we see that, when a small amount of error (condition '$\mathbf{P}_a\mathbf{P}_a$', see Appendix A) is added to the factors, all factors are retrieved by the three SCA–methods. With a intermediate amount of error (condition '$\mathbf{P}_b\mathbf{P}_b$'), factors are still almost always retrieved. However, when only weak factors are present (condition '$\mathbf{P}_c\mathbf{P}_c$'), SCA–S, SCA–P, and SCA–W recover only 71.0%, 69.4% and 62.4% of the true factors, respectively, and the differences between methods are considerable.

When half of the factors are strong and half of the factors are intermediate in the population (condition '$\mathbf{P}_d\mathbf{P}_d$'), factors are almost always retrieved. When half of the factors are strong and half of the factors are weak in the population (condition '$\mathbf{P}_e\mathbf{P}_e$'), factors are retrieved in about 90% of the cases, and SCA–S lags about 3% behind with the other two SCA–methods.

Finally, and most interestingly, when factors are strong in the first sample, and weak in the second sample (condition '$\mathbf{P}_a\mathbf{P}_c$') – remember that in this situation both samples are defined as coming from the same population, with as only difference a small amount of measurement error in the first sample and a large amount of measurement error in the second sample – both SCA–P and SCA–S retrieve exactly 100% of the factors, while SCA–W has a RR of only 87.8%. This is the main contribution to the PatternComb × Method interaction. A closer look at these results for

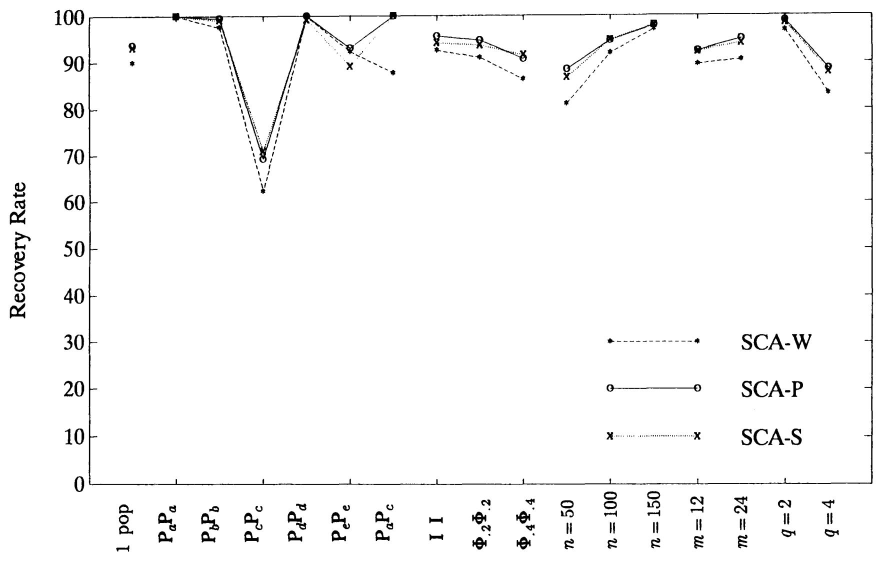


Figure 5.6 *Recovery Rates for the SCA–methods, averaged over each level of each independent variable, within the samples from one population*

SCA–W reveals that SCA–W always retrieves the factors from the first sample with strong factors ($\mathbf{P}_a$), but has a RR of only 75.6% for the factors from the second sample with weak factors ($\mathbf{P}_c$). Separating the '2 factors' from the '4 factors' condition within the condition '$\mathbf{P}_a\mathbf{P}_c$' shows that the RR in the second sample is 88.3% in the '2 factors' condition and 62.8% in the '4 factors' condition. Note that for SCA–S and SCA–P there is hardly any difference between these conditions because the RR is 100%. For PatternComb condition '$\mathbf{P}_c\mathbf{P}_c$', separating the RR's for the '2 factors' and '4 factors' condition gives acceptable RR's of 92.8%, 93.1% and 86.7% for SCA–S, SCA–P, and SCA–W, respectively, in the '2 factors' condition and a troublesome 49.3%, 45.7% and 38.1% for the respective methods in the '4 factors' condition. Together these results explain the significant interaction effect of PatternComb × NFac × Method (see Table 5.4).

There is a significant interaction effect of SampSize × Method.

SCA–W is more vulnerable to small sample sizes than SCA–P and SCA–S. Furthermore, one of the strongest effects is the interaction of PatternComb × SampSize × Method (see Table 5.4). For the conditions '$\mathbf{P}_c\mathbf{P}_c$' and '$\mathbf{P}_a\mathbf{P}_c$' of PatternComb, SCA–W has disproportionately low RR's at small sample size.

In total, the five effects mentioned here explain only a moderate 25.6% of the total variation, while the residual within subjects variation makes up 69.4% of the total variation.

A reason for the low RR's for SCA–W at small sample size can perhaps be found in the fact that, of the SCA–methods, SCA–W is the method that explains the most variance, and has been shown to have a positive sampling bias. This 'greed' for variance to explain may lead the method away from the underlying solution.

5.6.3.2 *Samples from two populations*

In Figure 5.7, the RR's are presented, summed over each condition of each independent variable within the condition 'samples from two populations', for the pattern matrix, found with SCA–W and SCA–P, and for the structure matrix, found with SCA–S. Overall, a highly significant method effect existed, explaining 18.6% of the total variation. SCA–P had a significantly higher RR (98.0%) than both SCA–W (95.1%) and SCA–S (86.2%), as opposed to the results for the '1 population' condition, where both SCA–P and SCA–S had higher RR's than SCA–W. SCA–W had a significantly higher RR than SCA–S. Summed over each condition of the independent variables SampSize, NVar and NFac separately (these are the independent variables that are known in practice), SCA–P always had the highest RR, followed by SCA–W and SCA–S, in that order.

There were strong interactions of PatternComb × Method and PhiComb × Method (see Table 5.4). Results for the six conditions of the independent variable PatternComb will not be discussed in detail, because each of the conditions had a very specific description. For instance, '$\mathbf{P}_c\mathbf{P}_g$' means 'weak factors in one sample' ($\mathbf{P}_c$) and 'factors with variables loading differently within each factor' ($\mathbf{P}_g$). To ease interpretation, it was

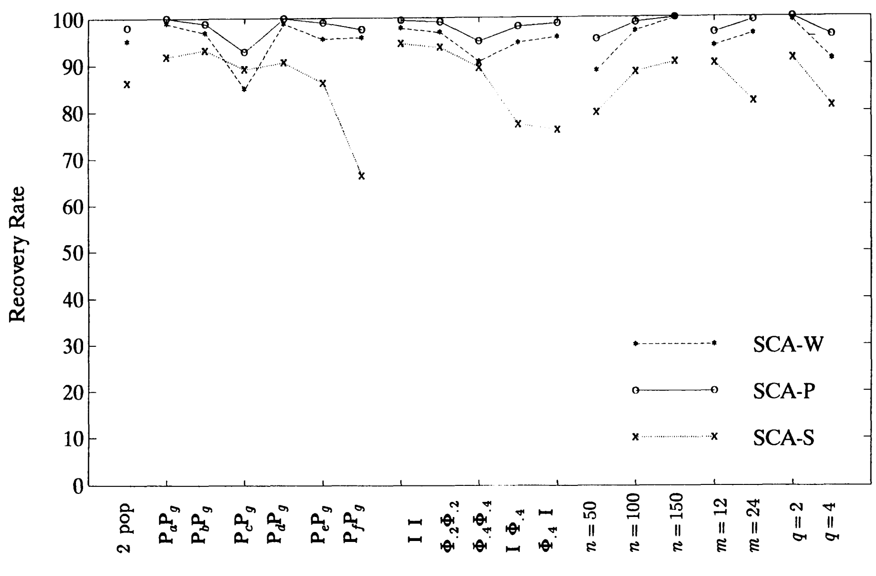


Figure 5.7 *Recovery Rates for the SCA–methods, averaged over each level of each independent variable, within the samples from two populations*

decided to form clusters of conditions combining pattern matrices and matrices Φ used, according to the CM–values of the true structure matrices, given in Table 5.1, and to compare results for these clusters. In order to keep the amount of error, added to the scores, equal over clusters, because this was a factor of great influence on the RR, for each cluster only those conditions were selected for which the mean noise–to–signal ratio used was 1.5. Only the conditions '$\mathbf{P}_b\mathbf{P}_g$', '$\mathbf{P}_e\mathbf{P}_g$' and '$\mathbf{P}_f\mathbf{P}_g$' satisfied this requirement. The fact that the other three conditions of PatternComb are not included here, does not mean that these conditions are dispensable. They were necessary for creating a broad range of possible pattern matrices within the condition 'samples from two populations'. The formed clusters were:

Cluster 1: $\mathrm{CM}_{s(tructure)}$ between .88 and .94; $\mathrm{CM}_{p(attern)}$ of .94 (n=720);
Cluster 2: CM_s between .80 and .87; CM_p of .94 (n=240);

Cluster 3: CM_s between .70 and .78; CM_p between .78 and .94 (n=720);
Cluster 4: CM_s of .64; CM_p of .78 (n=120).
The average values of the RR for these clusters are presented in Figure 5.8.

From Figure 5.8 it becomes clear that only SCA-S is sensitive to the CM-value of the true-matrices. For Clusters 3 and 4, the RR drops considerably. These results for SCA-S are not very surprising, because the true structure matrices are very different in the two populations in Cluster 4. Thus, it is impossible for SCA-S, which seeks one structure matrix that is the same for all samples, to retrieve both structure matrices. One could expect that SCA-S would better be able to retrieve the true pattern matrices in these situations, because the pattern matrices, found with SCA-S, are allowed to differ across groups. However, the average RR of the pattern matrices is only higher than the average RR of the structure matrix in Clusters 2 and 4, and lower in the other Clusters, as can be seen in Table 5.5. It is remarkable that the RR for SCA-P is as high as it is in Clusters 3 and 4, because the method only produces one pattern matrix. This pattern matrix apparently has high congruence values with both true pattern matrices in Clusters 3 and 4.

These results on differences between RR for pattern and structure led us to inspect the RR of the structure matrix for SCA-W and SCA-P. The RR of the structure matrix is higher than the RR of the pattern matrix in

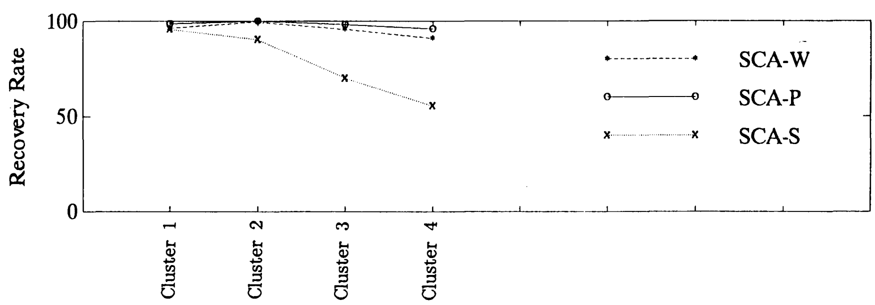


Figure 5.8 *Recovery Rates for the SCA-methods, in four clusters of samples with different (decreasing) CM values for the true structure matrices*

Table 5.5: *Recovery Rates of the pattern and structure matrices, found with the three SCA-methods in the four clusters (in %)*

	SCA–W		SCA–P		SCA–S	
	P	S	P	S	P	S
Cluster 1	96.1	95.8	98.7	96.4	91.5	95.6
Cluster 2	99.5	98.8	100.0	99.3	97.0	90.2
Cluster 3	95.6	96.2	98.1	97.0	66.1	69.8
Cluster 4	90.9	94.9	95.8	97.0	60.1	55.4

Cluster 3 for SCA–W and in Cluster 4 for both methods. So when the differences between the populations become larger, the structure matrix appears to remain closer to the correct factors than the pattern matrix does.

The strong interaction of PhiComb × Method was mainly caused by the low RR's for SCA–S in the conditions '$\mathbf{I}\ \mathbf{\Phi}_{.4}$' and '$\mathbf{\Phi}_{.4}\ \mathbf{I}$'. From Table 5.1 it can be seen that the CM of the true structure matrices is the lowest in the conditions '$\mathbf{I}\ \mathbf{\Phi}_{.4}$' and '$\mathbf{\Phi}_{.4}\ \mathbf{I}$', thus explaining this effect.

There also was a considerable interaction effect of NVar × Method (see Table 5.4). The RR of SCA–S was lower in the condition '24 variables' than in the condition '12 variables', while the RR's of SCA–W and SCA–P were higher in the condition '24 variables' than in the condition '12 variables' (see Figure 5.7). The effect of NVar, just mentioned, was especially different for each method in the PatternComb conditions '$\mathbf{P}_c\mathbf{P}_g$' and '$\mathbf{P}_f\mathbf{P}_g$', causing a strong 3-way interaction of PatternComb × NVar × Method. Worth noting in this respect is the enormous difference in RR for SCA–S within the condition '$\mathbf{P}_f\mathbf{P}_g$', between the conditions '12 variables' and '24 variables'. While in the '12 variables' condition, the RR for the method is 84.0%, in the '24 variables' condition, the RR is 48.6%, a difference of 35.4%. For comparison, within the other conditions of PatternComb, when comparing the '12 variables' with '24 variables' condition, the difference in RR for SCA–S varies between –5.5% and +2.9% (the only positive difference was found in condition '$\mathbf{P}_c\mathbf{P}_g$'). Finally, there was a moderate 3-way interaction of

PhiComb × NFac × Method: PhiComb strongly influences the success of SCA-S. This influence is caused by the low CM of the true structure matrices for two conditions of PhiComb. The negative effect on the RR of SCA-S in these two conditions of PhiComb is stronger when four factors are present than when two factors are present. This influence is again caused by the lower CM of the true structure matrices when four factors are present than when two factors are present (see Table 5.1).

In total, the six effects mentioned here explain 60.0% of the total within subjects variation.

In answer to Question 3a, it is concluded that, when samples are indeed coming from one population (one cannot be certain of this in practice), SCA-P and SCA-S are best capable of recovering the underlying factors, although the differences in success rates between the three SCA-methods are small. When samples are coming from two different populations, SCA-P is still best capable of retrieving the underlying factors, followed by SCA-W and SCA-S, in that order. The capability of SCA-S to retrieve the underlying factors is considerably impaired when the two populations differ markedly. This could be seen as a shortcoming, but according to the author it is actually an asset of the method, because the populations indeed differ, so the method should not be capable of retrieving common factors. It remains a problem, however, how this population difference can be detected in practice (See Section 5.6.5).

In answer to Question 3b, it can be concluded that the independent variable PatternComb has the largest influence on the relative success of each method in recovering underlying factors. It is, furthermore, interesting that for 'samples from two populations', the choice of PhiComb strongly influences the success of SCA-S. This influence is caused by the low CM of the true structure matrices for two conditions of PhiComb. For the 'samples from two populations, the number of variables influences the relative success of each method. While SCA-W and SCA-P have higher recovery rates when 24 variables are present than when 12 variables are present, the opposite is the case for SCA-S. This coincides with the results for the measure QDA.

5.6.4 Finding the SCA-method with the best recovery of correlations between factors

Question 4a: "Which of the SCA–methods can be judged as best capable of recovering the correlations between factors?"
Question 4b: "Are there interesting interaction effects between independent variables and method of analysis for the recovery of correlations between factors?"

The average DFC values (see Section 4.4.2) for samples from one population were .115 for both SCA–W and SCA–P and it was .122 for SCA–S. Note that the DFC values indicate how badly factor correlations are recovered, hence better recovery is indicated by lower values. For the samples from two populations, the average DFC values were .104, .106 and .217 for SCA–W, SCA–P and SCA–S respectively. The different results for samples from one and two populations for SCA–S and the similar results for samples from one and two populations for SCA–W and SCA–P asked for separate analysis of these two conditions.

To get an overview of the size of the main effect of Method and its interactions with one or two between subject factors, the associated sums of squares for the success criterion DFC in the conditions 'samples from one population' and 'samples from two populations', are presented in Table 5.6, where significant effects at α=.001 are again marked by asterisks.

For all conditions of the independent variables within the condition 'samples from one population', the differences between the three SCA–methods were small, although significant, except for NFac. As can be seen from Table 5.6, the total within subjects variation was very small in the condition 'samples from one population', as compared with the condition 'samples from two populations'. Even when correcting for the larger number of cases in the latter condition, the variation in the condition 'samples from two populations' was still 17 times the variation in the condition 'samples from one population'. The largest effect in the condition 'samples from one population' was PhiComb × Method. The mean

Table 5.6: *Sums of Squares for the measure DFC from the ANOVA analysis of samples from one population* ($n = 2160$) *and samples from two populations* ($n = 3600$)

	one population	two populations
Total within subjects	2.44	70.41
Method	.07*	30.15*
PatternComb × Method	.08*	.61*
PhiComb × Method	.40*	26.28*
SampSize × Method	.17*	.07*
NVar × Method	.09*	.26*
NFac × Method	.00	.20*
PatternComb × PhiComb × Method	.20*	1.94*
PatternComb × SampSize × Method	.04*	.12*
PatternComb × NVar × Method	.01	.20*
PatternComb × NFac × Method	.02*	.10*
PhiComb × SampSize × Method	.03*	.54*
PhiComb × NVar × Method	.01	.18*
PhiComb × NFac × Method	.02*	.12*
SampSize × NVar × Method	.05*	.00
SampSize × NFac × Method	.00	.02
NVar × NFac × Method	.00	.17*
Residual within subjects	1.25	9.43

Method = Method used, PatternComb = Used Pattern Combinations, PhiComb = Used Combinations of matrices Φ, *NFac = Number of factors, NVar = Number of variables. * = significant at* $\alpha = .001$.

values of the DFC measure for SCA–W, SCA–P and SCA–S, were respectively .08, .09, and .11 in PhiComb condition '**I I**', .11, .11, and .12 in PhiComb condition '$\mathbf{\Phi}_{.2}\mathbf{\Phi}_{.2}$', and .16, .15, and .14 in PhiComb condition '$\mathbf{\Phi}_{.4}\mathbf{\Phi}_{.4}$', illustrating that differences in differences between the methods were indeed small. Therefore, the differences, found in the condition 'samples from one population', were deemed not interesting.

Within the condition 'samples from two populations' there was a sizable Method effect. The mean DFC value was larger for SCA–S than for SCA–W and SCA–P. By far the strongest interaction was of PhiComb × Method. In Figure 5.9, DFC values for the five conditions of the independent variable PhiComb, within the condition 'samples from two populations', are presented for each method.

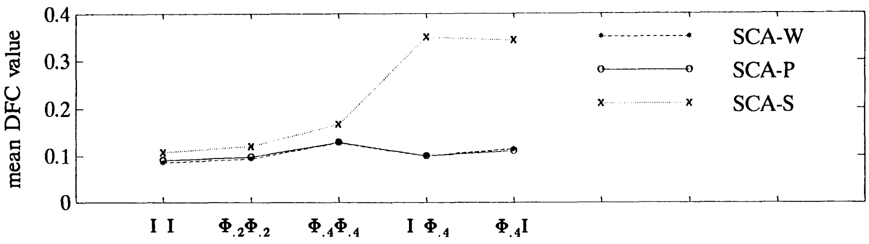


Figure 5.9 *Mean DFC values for the SCA-methods, averaged over the five levels of the independent variable PhiComb, within the samples from two populations*

From Figure 5.9, it can be seen that in the conditions '**I** $\mathbf{\Phi}_{.4}$' and '$\mathbf{\Phi}_{.4}$ **I**' SCA-S has a much larger DFC value than SCA-W and SCA-P. Recall that for these two conditions of PhiComb the CM of the true structure matrices was the lowest, and the RR of the structure matrix by SCA-S was also the lowest. Together, the effects of Method and the interaction of PhiComb × Method explain 80.1% of the total variation between subjects.

Answers 4a and 4b: For the samples from one population, there were no large differences between the SCA-methods in recovering the correlations between the factors. When samples come from two different populations, SCA-W and SCA-P give correlations closer to the correct values than SCA-S. This difference is mainly found in the condition where the samples from two populations are based on different factor correlation matrices.

5.6.5 Discerning the 'samples from one population' from the 'samples from two populations' condition

Question 5: "Can samples from one population be distinguished from samples from two populations?"

A discriminant analysis was performed to see to what extent we can correctly classify each data set (consisting of two samples) in the category 'samples from one population' or in the category 'samples from two populations'. A subset of data sets was selected for inclusion in the analysis. The selection was made on the basis of two requirements. First, the overall amount of error in the data sets had to be the same in all data sets, because otherwise the (in practice unobservable) amount of error in the data could be the strongest aspect influencing the discriminant function and the desired discriminant function could be blurred. Secondly, the selected PatternComb conditions on which the samples from two populations were based had to be 'considerably' different, that is, only those data sets were selected for which the CM of the true pattern matrices was lower than .85 (see Table 5.1). On the basis of these two requirements, the 1080 data sets based on the PatternComb conditions '$\mathbf{P}_b\mathbf{P}_b$', '$\mathbf{P}_e\mathbf{P}_e$' and '$\mathbf{P}_a\mathbf{P}_c$' were selected to represent samples from one population, and the 600 data sets based on the PatternComb condition '$\mathbf{P}_f\mathbf{P}_g$' were selected to represent samples from two populations.

In the discriminant analysis, the following seven predictor variables were used: SampSize, NVar, NFac and the percentages of variance explained by the three SCA-methods and PCA-sep ($EV_{SCA\text{-}W}$, $EV_{SCA\text{-}P}$, $EV_{SCA\text{-}S}$ and $EV_{PCA\text{-}sep}$). These seven predictor variables are all the variables that are observed (or can be observed) in practice.

In a first analysis, all of the predictors mentioned above were included. The overall percentage of correct classifications was 92.2%. The percentage of correct classifications was 95.8% for samples from one population (PatternComb conditions '$\mathbf{P}_b\mathbf{P}_b$', '$\mathbf{P}_e\mathbf{P}_e$' and '$\mathbf{P}_a\mathbf{P}_c$') and 85.7% for samples from two populations (PatternComb condition '$\mathbf{P}_f\mathbf{P}_g$'). The predictor variable 'percentage of explained variance of SCA-P' did not pass the tolerance test and was excluded from the discriminant function. The tolerance is a measure of the degree of linear association between the independent variables, implemented in SPSS (Norušis, 1990): "For the i^{th} independent variable, it is [calculated as] $1-R_i^2$, where R_i^2 is the squared multiple correlation coefficient when the i^{th} independent variable is considered the dependent variable and the regression equation between it and the other independent variables is calculated. Small values for the tolerance indicate that the i^{th} independent variable is almost a linear combination of the other independent variables. Variables with small tolerances (by default, less than .001) are not permitted to enter the analysis."

The predictor variable 'percentage of explained variance of SCA–W' had a very small standardized coefficient (–.15), compared to that for the explained variances of PCA–sep (–3.00) and SCA–S (4.04). Because of this, a second discriminant analysis was done, leaving out the explained variances of SCA–P and SCA–W as predictor variables. This resulted in a discriminant function that correctly classified the cases about just as often (overall percentage of correct classifications was 92.3%; 95.8% for samples from one population and 85.8% for samples from two populations). Finally, a third discriminant analysis was done to see whether inclusion of the predictor variables SampSize, NVar and NFac was necessary, so only the explained variance of PCA–sep and SCA–S were used as predictor variables. The percentages of correct classifications was now 86.8% for the samples from one population and 77.2% for the samples from two populations (overall 83.3%). The loss in correct classifications in the third analysis was considered too large, and therefore the discriminant function from the second analysis was considered most useful for classifying data sets as coming from one population or from two 'considerably' different populations.

To crossvalidate the results from the discriminant analysis, the selection of cases used for the discriminant analysis (PatternComb conditions '$\mathbf{P}_b\mathbf{P}_b$', '$\mathbf{P}_e\mathbf{P}_e$', '$\mathbf{P}_a\mathbf{P}_c$' and '$\mathbf{P}_f\mathbf{P}_g$') was split in two. That is, the ten replications within each condition were evenly divided over the two halves. On each of the halves, denoted as A and B, another discriminant analysis was performed and the two discriminant functions thus found, denoted as D_A and D_B, were put to the test on the other half. The percentages of correct classifications using D_A on A were 97.0% for samples from one population and 84.7% for samples from two populations. Using D_B on A gave correct classifications of 98.0% and 83.3%, respectively. (Specifically, in total 779 cases were correctly classified in the crossvalidation of D_B on A, against 778 when using D_A on A. This perhaps counterintuitive improvement is not disturbing, because the optimized discriminant function does not necessarily gives the best classification; it only does so when the distribution of the groups discriminant scores is normal.) The percentages of correct classifications using D_B on B were 94.8% for samples from one population and 85.0% for samples from two populations. Using D_A on B gave correct classifications of 90.0% and 89.3%, respectively. The percentages of correct classifications show good crossvalidation of the results. The coefficients in the two discriminant functions were very similar, differing .02 at most, except for the predictor variable NFac, where the unstandardized coefficient was −.36 in D_A and −.26 in D_B. This predictor variable also had the smallest standardized discriminant function coefficient. From this crossvalidation, it was concluded that the discriminant weights were very stable, except for the weights for predictor variable NFac. For this reason, it was decided do a final discriminant analysis, in which that predictor variable was left out. To further simplify the thus found discriminant function, a simple version of it was tested, in which only the first three decimals from the complete discriminant function were used. Denoting the percentage of explained variance by EV, the simple discriminant function can be written as

$$d(EV_S,EV_A,NVar,SS) = .469EV_S-.431EV_A+.136NVar-.016SS+.453, \qquad (5.1)$$

in which $EV_S=EV_{SCA-S}$, $EV_A=EV_{PCA-sep}$ and SS=SampSize. When $d>0$, the discriminant function indicates that the samples are *probably* from one population. When $d<0$, it indicates that the samples are *probably* from two different populations.

As a second validation, the complete discriminant function (i.e., function (5.1) with 7 decimals instead of three) and the simple discriminant function were used on all data sets, categorized with respect to the twelve different PatternComb conditions used in this experiment. The percentages of correct classifications for each PatternComb condition are given in Table 5.7.

From Table 5.7, it can be seen that there are only small differences

Table 5.7: *Percentages of correct classifications for the two discriminant functions for each condition of PatternComb*

	complete function	*simple function*
One–population data		
$\mathbf{P}_a\mathbf{P}_a$	95.6	95.6
$\mathbf{P}_b\mathbf{P}_b$	96.9*	96.9*
$\mathbf{P}_c\mathbf{P}_c$	79.4	82.2
$\mathbf{P}_d\mathbf{P}_d$	97.8	97.8
$\mathbf{P}_e\mathbf{P}_e$	98.6*	98.6*
$\mathbf{P}_a\mathbf{P}_c$	95.0*	95.3*
Two–or–more–populations data		
$\mathbf{P}_a\mathbf{P}_g$	55.2	52.0
$\mathbf{P}_b\mathbf{P}_g$	27.7	24.2
$\mathbf{P}_c\mathbf{P}_g$	18.7	15.8
$\mathbf{P}_d\mathbf{P}_g$	43.3	40.2
$\mathbf{P}_e\mathbf{P}_g$	26.8	21.5
$\mathbf{P}_f\mathbf{P}_g$	80.5*	78.3*

* *indicates conditions used in the discriminant analysis*

between the two discriminant functions, thus suggesting that the simple function can be used in the sequel. However, some at first sight disturbing results were found. In the crossvalidation cases, the discriminant function only gives good results for the PatternComb conditions '$\mathbf{P}_a\mathbf{P}_a$' and '$\mathbf{P}_d\mathbf{P}_d$'. For all other PatternComb conditions, the discriminant function classifies the cases wrongly for more than 50% of the data sets. The question to ask at this point is: "How serious are these misclassifications?"

The fact that samples from the PatternComb condition '$\mathbf{P}_c\mathbf{P}_c$' are classified as coming from two populations in 53% of the cases is not disturbing, because these samples are very poor anyway (they contain a very large amount of error), and thus a researcher is obliged to proceed with caution. The fact that samples based on the PatternComb conditions '$\mathbf{P}_a\mathbf{P}_g$' to '$\mathbf{P}_e\mathbf{P}_g$' are often classified as coming from one population is also not very disturbing, because from the CM of the true pattern matrices (see Table 5.1) it can be seen that the samples come from different, but not 'considerably' different populations, and thus, in practice, it is acceptable to view the samples as coming from one population. Finally, with the present discriminant function an easy rule of thumb can be given for discerning samples from one population with samples from two populations. If the discriminant value indicates that the samples come from two different populations, it is very likely that this is indeed the case.

Answer 5: By using discriminant function (5.1), it can be determined reasonably reliably whether or not samples come from two populations with distinctly different underlying patterns.

5.6.6 Recovery of factors for specific cases of samples from one population

To investigate the ability of the SCA–methods to cope with different variations in strength of factors, several separate comparisons were made.

Question 6a: "Which SCA–method gives the highest RR, when comparing samples taken from one population, with only strong factors (condition '$\mathbf{P}_a\mathbf{P}_a$''), only intermediate factors (condition '$\mathbf{P}_b\mathbf{P}_b$''), and only weak factors (condition '$\mathbf{P}_c\mathbf{P}_c$'')?"

In Table 5.8, the RR's for the SCA–methods are given for four conditions of PatternComb (within each PatternComb condition RR's are averaged over all other independent variables), describing samples from one population. The first three conditions ('$\mathbf{P}_a\mathbf{P}_a$'', '$\mathbf{P}_b\mathbf{P}_b$'' and '$\mathbf{P}_c\mathbf{P}_c$'') are also depicted in Figure 5.6. For the condition in which factors of different strength were present in the two samples ('$\mathbf{P}_a\mathbf{P}_c$''), RR's for the factors of different strength are given separately.

There was a significant interaction effect between PatternComb and Method (see Figure 5.6). Comparing two methods at a time, it appeared that SCA–W showed a significantly larger drop in RR when going from pattern '$\mathbf{P}_a\mathbf{P}_a$'' to '$\mathbf{P}_c\mathbf{P}_c$'', than both SCA–S and SCA–P. The difference

Table 5.8: *RR's for the SCA-methods for the conditions of PatternComb describing samples from one population, with RR's separated for factors of different strength*

	SCA-W		SCA-P		SCA-S	
$\mathbf{P}_a\mathbf{P}_a$ $(=s)$	1.00		1.00		1.00	
$\mathbf{P}_b\mathbf{P}_b$ $(=i)$	.98		.99		.99	
$\mathbf{P}_c\mathbf{P}_c$ $(=w)$	.62		.69		.71	
Fact.str.	*s*	*w*	*s*	*w*	*s*	*w*
$\mathbf{P}_a\mathbf{P}_c$	1.00	.76	1.00	1.00	1.00	1.00

s=strong (loading=.89), i=intermediate strength (.55), w=weak (.37)

between SCA-S and SCA-P was not significant at α=.001 (p=.048).

Answer 6a: When large amounts of error are present, SCA-P and SCA-S give the best results on the RR measure. When small or intermediate amounts of error are present, all methods have almost complete recovery of factors.

Question 6b: "Are, for the SCA-methods, intermediate factors and weak factors blurred by strong factors, or vice versa?"

For this question, the influence of factors of different strengths within a sample was investigated. To clarify the comparison, keep the following situation in mind. One has scores on a set of 24 variables, pertaining to 4 factors: 2 strong factors and 2 intermediate or weak factors (as in $\mathbf{P}_d$ and $\mathbf{P}_e$). Each factor is defined by nonzero loadings on 6 variables. If one would analyze the 12 variables making up the intermediate or weak factors separately (as in data based on $\mathbf{P}_b$ and $\mathbf{P}_c$), one would like to see the same results for these factors as when analyzing the complete data set with 24 variables. To see whether this requirement is fulfilled by the SCA-methods, the '12 variables and 2 factors' conditions within the conditions '$\mathbf{P}_b\mathbf{P}_b$' and '$\mathbf{P}_c\mathbf{P}_c$' were compared with the '24 variables and 4 factors' conditions within the conditions '$\mathbf{P}_d\mathbf{P}_d$' and '$\mathbf{P}_e\mathbf{P}_e$', respectively (see Table 5.9). Also comparisons of condition '$\mathbf{P}_a\mathbf{P}_a$' and the conditions '$\mathbf{P}_d\mathbf{P}_d$' and '$\mathbf{P}_e\mathbf{P}_e$' were made. In these comparisons the variable–factor ratio is held constant.

Comparing the RR of the strong factors in condition '$\mathbf{P}_a\mathbf{P}_a$' with '$\mathbf{P}_d\mathbf{P}_d$' and '$\mathbf{P}_e\mathbf{P}_e$', we see that the RR for the strong factors drops for SCA–W and SCA–S, when weak factors are added to the analysis. SCA–P is not troubled by this effect. Comparing the RR of the factors of intermediate strength in condition '$\mathbf{P}_b\mathbf{P}_b$' with '$\mathbf{P}_d\mathbf{P}_d$', we see no effects, because recovery of the factors is 100%, except for a few factors not recovered by SCA–S (10 out of 360). Comparing the RR of the weak factors in condition '$\mathbf{P}_c\mathbf{P}_c$' with '$\mathbf{P}_e\mathbf{P}_e$', we see that the RR for SCA–P drops from .88 to .77, while it drops very little for both SCA–W and SCA–S. So, in this condition, SCA–P seems to favor the recovery of the strong factors at the cost of the recovery of the weak factors, and SCA–W and SCA–S seem to favor the recovery of the weak factors at the cost of the recovery of the strong factors.

Answer 6b: In the condition where the variable–factor ratio is held equal, for SCA–W and SCA–S, it was found that the presence of weak factors, within each sample, decreases the recovery of strong factors. For SCA–P, the presence of strong factors, within each sample, decreases the recovery of weak factors.

Table 5.9: *RR's for the SCA-methods for the* 12 *variables/*2 *factors conditions of levels* '$\mathbf{P}_a\mathbf{P}_a$', '$\mathbf{P}_b\mathbf{P}_b$' *and* '$\mathbf{P}_c\mathbf{P}_c$', *and for the* 24 *variables/*4 *factors conditions of levels* '$\mathbf{P}_d\mathbf{P}_d$' *and* '$\mathbf{P}_e\mathbf{P}_e$' (*number of factors per cell is* $n=360$)

	SCA–W		SCA–P		SCA–S	
$\mathbf{P}_a\mathbf{P}_a$ (=*s*)	1.00		1.00		1.00	
$\mathbf{P}_b\mathbf{P}_b$ (=*i*)	1.00		1.00		1.00	
$\mathbf{P}_c\mathbf{P}_c$ (=*w*)	.84		.88		.90	
Fact.str.	*s*	*i*	*s*	*i*	*s*	*i*
$\mathbf{P}_d\mathbf{P}_d$	1.00	1.00	1.00	1.00	.98	.97
Fact.str.	*s*	*w*	*s*	*w*	*s*	*w*
$\mathbf{P}_e\mathbf{P}_e$	.90	.83	1.00	.77	.81	.87

s=strong (loading=.89), *i=intermediate strength* (.55), *w=weak* (.37)

Question 6c: "Is the RR for the three SCA–methods affected by the existence of small measurement error in one sample and large measurement error in another sample (condition '$\mathbf{P}_a\mathbf{P}_c$'), and which SCA–method performs best under this condition?"

We return to Table 5.6 one more time to compare the RR for the SCA–methods in the conditions '$\mathbf{P}_a\mathbf{P}_a$' and '$\mathbf{P}_c\mathbf{P}_c$' with '$\mathbf{P}_a\mathbf{P}_c$', to see whether RR's were affected by the existence of small measurement error in one sample and large measurement error in another sample. From Table 5.8, we see that the strong factors were always recovered. Furthermore, for all SCA–methods the recovery of the weak factors is enhanced when one of the samples contains strong factors (scores with little measurement error). SCA–W, however, benefits only slightly from the presence of a sample with strong factors.

Answer 6c: The methods SCA–P and SCA–S are fully capable of retrieving weak factors in one sample, when the same factors in the other sample are strong. SCA–W does not have this capability.

5.7 CONCLUSION

Of the four dimension indicators tested, the measure QDA gave the best overall results for indicating the correct number of components for all methods of analysis. For this reason, the measure QDA will be used as the success criterion for dimension indication in the remainder of this study.

SCA–P gives the best results for all dimension indicators. The measure QDA is the dimension indicator that can best be used when applying SCA–W, SCA–P and PCA–sep, and the measure QA is the dimension indicator that can best be used when applying SCA–S. Surprisingly, differences between the dimension indicators are not much influenced by sample size, although all dimension indicators have higher success rates for larger sample sizes. The effect of the number of factors present is much stronger. Four components are much harder to be correctly indicated by the measures PA, QA and QDA, while the measure KA1 has more trouble correctly indicating two components. When four factors are present, all four dimension indicators perform about equally well for all methods, although the measure QDA still has the highest success rate.

Using the measure QDA, SCA–W and SCA–P (including the rotations, as described in Chapter 2) were most successful in indicating the correct number of components. SCA–P is the method recommended, when the correct number of components has to be determined solely from the data, that is, when the number of components can not be decided on on theoretical grounds, because this method takes the shortest time to execute.

When samples are coming from one population, SCA–P and SCA–S are somewhat better capable of recovering the underlying factors than SCA–W. This method especially falls behind when only weak factors are present and when there are weak factors in one sample, when the same factors in the other sample are strong. SCA–W fails to recover factors in situations where, compared to the other SCA-methods, it can gain a relatively large percentage of variance. These situations are at small sample sizes (see Section 5.6.3.1) and when there are weak components present (large amounts of error). SCA–W seems to show the property of encompassing a relatively large portion of 'error' in its components. It must be noted, however, that mostly the differences in success rates between the three SCA-methods are small. The lower recovery rate of SCA–W in the conditions mentioned above is especially apparent at small sample size (n=50). For SCA–P, the presence of strong factors within each sample decreases the recovery of weak factors. For SCA–W and SCA–S, the presence of weak factors within each sample decreases the recovery of strong factors. When samples are coming from two different populations, SCA–P is best capable of retrieving the underlying factors, followed by SCA–W and SCA–S, in that order.

The results on retrieving the correlations between the factors reveal that for samples from one population, all three SCA-methods perform about equally good. When samples come from two 'considerably' different populations, SCA–W and SCA–P give correlations closer to the correct values than SCA–S.

Overall, SCA–P has the best results on dimension indication and on retrieving factors and factor correlations and is therefore recommended. SCA–W performs somewhat more poorly, and SCA–S performs about just as well as SCA–P when the samples come from one population, and performs worse than SCA–P when the samples come from two populations.

If it is desired, when doing simultaneous components analyses, to determine reasonably reliably whether samples come from two populations with distinctly different underlying patterns or not, we would advise to use discriminant function (5.1). It is, however, advised to use the discriminant function only when the number of variables and the sample size used are within the range of the values used in the determination of the discriminant function. Furthermore, it might be wise to use both SCA–P and SCA–S, and see whether the two methods lead to different interpretations of the components found. If so, this is an additional indication that there are considerable differences between the populations from which the samples were drawn, because in this case the solutions from SCA–S tend to differ considerably from the solutions from SCA–P (where the latter still adequately recover underlying factors, whereas the former to a large extent fail to do so).