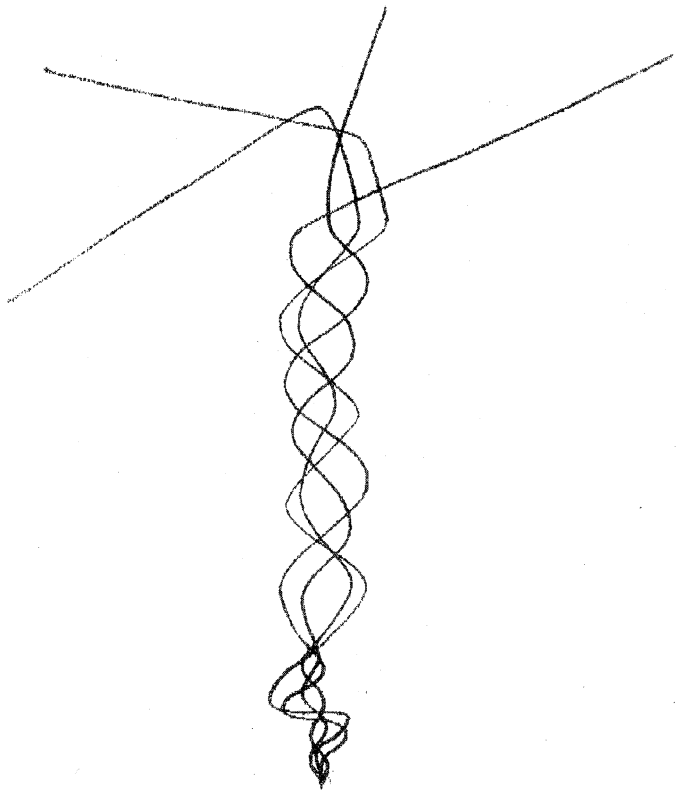
SIMULTANEOUS COMPONENT AND FACTOR ANALYSIS METHODS FOR TWO OR MORE GROUPS: A COMPARATIVE STUDY

Jan Niesing



DSWO PRESS

SIMULTANEOUS COMPONENT AND FACTOR ANALYSIS METHODS FOR TWO OR MORE GROUPS: A COMPARATIVE STUDY

M&T SERIES 30

Editorial Board:

Prof.dr. J.M.F. ten Berge
Prof.dr. W.J. Heiser
Prof.dr. L.J.Th. van der Kamp
Prof.dr. J. de Leeuw

Technical Editor: L. Delvaux

Earlier publications in this series:

Jacqueline Meulman, Homogeneity Analysis of Incomplete Data.
M&T series 1, 1982
Pieter M. Kroonenberg, Three-mode Principle Component Analysis.
M&T series 2, 1983 (reprint 1989)
Jan de Leeuw, Canonical Analysis of Categorical Data.
M&T series, 1984
Ronald A. Visser, Analysis of Longitudinal Data in Behavioural and Social Research.
M&T series 4, 1985
John P. van de Geer, Introduction to Linear Multivariate Data Analysis.
M&T series 5, volume 1 & 2, 1986
Jacqueline Meulman, A Distance Approach to Nonlinear Multivariate Analysis.
M&T series 6, 1986
Jan de Leeuw, Willem Heiser, Jacqueline Meulman, Frank Critchley (editors), Multidimensional Data Analysis.
M&T series 7, 1987
Peter G.M. van der Heijden, Correspondence Analysis of Longitudinal Categorical Data.
M&T series 8, 1987
Jan van Rijckevorsel, The Application of Fuzzy Coding and Horseshoes in Multiple Correspondence Analysis.
M&T series 9, 1987
Abby Israëls, Eigenvalue Techniques for Qualitative Data.
M&T series 10, 1987
Eeke van der Burg, Nonlinear Canonical Correlation and Some Related Techniques.
M&T series 11, 1988
Kees van Montfort, Estimating in Structural Models with Non-normal Distributed Variables: Some Alternative Approaches.
M&T series 12, 1989
Jan T.A. Koster, Mathematical Aspects of Multiple Correspondence Analysis for Ordinal Variables.
M&T series 13, 1989
Catrien C.J.H. Bijleveld, Exploratory Linear Dynamic Systems Analysis.
M&T series 14, 1989
Henk A.L. Kiers, Three-way Methods for the Analysis of Qualitative and Quantitative Two-way Data.
M&T series 15, 1989
Stef van Buuren, Optimal Scaling of Time Series.
M&T series 16, 1990

Rien van der Leeden, Reduced Rank Regression with Structured Residuals.
M&T series 17, 1990
Rian van Blokland-Vogelesang, Unfolding and Group Consensus Ranking for Individual Preferences.
M&T series 18, 1991
Jacques J.F. Commandeur, Matching Configurations.
M&T series 19, 1991
Herbert Hoijtink, PARELLA, Measurement of Latent Traits by Proximity Items.
M&T series 20, 1991
Wendy J. Post, Nonparametric Unfolding Models, a Latent Trait Approach.
M&T series 21, 1992
Ivo A. van der Lans, Nonlinear Multivariate Analysis for Multiattribute Preference Data.
M&T series 22, 1992
Wim P. Krijnen, The Analysis of Three-Way Arrays by Constrained Parafac Methods.
M&T series 23, 1992
Marijtje A.J. van Duijn, Model for Repeated Counts
M&T series 24, 1993
Jos M.F. ten Berge, Least Squares Optimization in Multivariate Analysis.
M&T series 25, 1993
Patrick J.F. Groenen, The Majorization Approach to Multidimensional Scaling: Some Problems and Extensions.
M&T series 26, 1993
Peter Verboon, A Robust Approach to Nonlinear Multivariate Analysis.
M&T series 27, 1994
Monica Th. Markus, Bootstrap Confidence Regions in Non-Linear Multivariate Analysis.
M&T series 28, 1994
Frits Meijerink, A Nonlinear Structural Relations Model.
M&T series 29, 1995
Jan Niesing, Simultaneous Component and Factor Analysis Methods for Two or More Groups: A Comparative Study.
M&T series 30, 1997

Available form DSWO Press
Leiden University
P.O. Box 9555
2300 RB Leiden
The Netherlands
Tel. (071) 5273795 / 5273794
Fax (071) 5273619

SIMULTANEOUS COMPONENT AND FACTOR ANALYSIS METHODS FOR TWO OR MORE GROUPS: A COMPARATIVE STUDY

Jan Niesing

DEPARTMENT OF PSYCHOLOGIE
University of Groningen

1997 DSWO Press, Leiden University, The Netherlands

© 1997 DSWO Press, Leiden University, The Netherlands
All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without prior permission of the publisher.

© Cover design: J. Niesing and Infofilm, Roger Busschots, Leiden
Printed by 'HAVEKA b.v.' and by
Printing office 'Karstens drukkers/designers', Leiden

CIP-Data Koninklijke Bibliotheek
Subject Headings: Principle Component Analysis, Factor Analysis, Cross Sectional Data Analysis
ISSN 0928-1088-30
ISBN 90-6695-134-6

ACKNOWLEDGEMENT

The past four years I have worked at my Ph.D. thesis with great pleasure. Many people have helped me with me and my study and the time for a grateful "thank you" has finally come.

First of all, I would like to thank Jos ten Berge and Henk Kiers for their intense coaching, their sharp and always useful comments, and making sure that this wild goose would keep focused on the study instead of flying around without direction (which is also fun, but does not get you anywhere). Furthermore they kept me from being too wordy, otherwise you would be holding only the first volume in a series of books about my research instead of the shortened version you are holding now.

For advices on the use of factor analysis methods and the analyses of the results from the experiments conducted I would like to thank Anne Boomsma and Tom Snijders, respectively. For reading the final manuscript I would like to thank the members of the review committee, prof. dr. G.J. Mellenbergh, prof. dr. W. Molenaar and prof. dr. W.J. Heiser.

A lot of pleasure came from the enjoyable atmosphere on the fourth floor of the Heymans building, where the department of Personality Psychology and the department of Statistics and Measurement Theory are located. Of the fourth floor colleagues I would especially like to mention my ex-room mates Francis Tuerlinckx, Jeroen Heres, and, last but not least, Gertrude Smit, with whom I shared my good and bad times.

As regards the graduation ceremony, I am very grateful to Wander Jager and Roy Bradwolff for assisting me and to get things organized.

Finally, I would like to thank my brother Wim and my parents Alie en Henk Niesing, for their strong and loving support throughout my study.

CONTENTS

CHAPTER 1 GENERALIZED COMPONENTS ANALYSIS

1.1	Introduction	1
1.2	Principal Components Analysis	2
1.3	Comparing the Varimax rotated components	5
1.4	Applying the Generalized Procrustes Rotation	6
1.5	Methods for simultaneous analysis in two or more groups, used in the present study	7
1.6	Goals of the present study	9
1.7	Other methods for the simultaneous analysis in two or more groups: IDIOSCAL, PARAFAC2 and INDSCAL	10
1.7.1	Multidimensional Scaling	11
1.7.2	IDIOSCAL	11
1.7.3	PARAFAC2	12
1.7.4	INDSCAL	14
1.7.5	Reasons for not selecting the MDS methods	15
1.8	Discarding the Partial Common Principal Components method	15
1.8.1	Partial Common Principal Components (PCPC)	15
1.8.2	Reasons for discarding PCPC	16
1.9	Factor Analysis versus Components Analysis in the one group situation	18
1.10	The organization of the present study	21

CHAPTER 2 THE METHODS UNDER INVESTIGATION

2.1	Introduction	23

2.2	Simultaneous Components Analysis methods	23
2.2.1	SCA–W	24
2.2.2	SCA–P	25
2.2.3	SCA–S	26
2.3	A Simultaneous Factor Analysis method	26
2.3.1	Simultaneous Factor Analysis in Several Populations, using a Maximum Likelihood estimation procedure (SIFASP–ML)	26
2.3.2	Identification of parameters	28
2.4	Rotations used to facilitate interpretation of the solutions, found with the methods of analysis	29
2.4.1	HKIC Rotation of the SCA–P and the SIFASP–ML solution	30
2.4.2	Hakstian's Rotation of the SCA–W solution	31
2.4.3	Simple Structure Rotation of the SCA–S solution	31
2.5	Measures of fit for the SCA methods and SIFASP–ML	32
2.5.1	Explained variances of the SCA–W solutions	32
2.5.2	Explained variances of the SCA–P solutions	33
2.5.3	Explained variances of the SCA–S solutions	33
2.5.4	The Chi–Square measure for SIFASP–ML	34
2.6	First comparison of the methods	34
2.6.1	Comparisons of PCA–sep, SCA–W, SCA–P and SCA–S	34
2.6.2	Comparison of the SCA–methods with SIFASP–ML	35

CHAPTER 3 CONSTRUCTING DATA WITH KNOWN FACTOR STRUCTURE

3.1	Introduction	37
3.1.1	Terminology	37
3.2	Description of the data construction	38
3.2.1	Choosing a true pattern matrix $\mathbf{P}_{tk}$ and defining a true structure matrix $\mathbf{S}_{tk}$ when simulating a correlation or a covariance matrix	39
3.2.2	The non–uniqueness of the true weights matrix $\mathbf{W}_{tk}$	41
3.3	Definition of data categories	42
3.3.1	Sampling from one population	43

3.3.2 Sampling from two or more populations 43
3.3.3 Explained variances in the populations 44
3.4 A measure for the similarity of populations 46
3.4.1 The Congruence Measure 46

CHAPTER 4 INDEPENDENT VARIABLES, DIMENSION INDICATORS AND SUCCESS CRITERIA

4.1 Manipulations 49
4.1.1 Independent Variables 49
4.1.2 The true matrices underlying the constructed data 51
4.2 Dimension indicators for the SCA–methods 53
4.2.1 Additional value greater than one (KA1) 53
4.2.2 Parallel Analysis (PA) 54
4.2.3 The Quotient of Additional values (QA) 55
4.2.4 The Quotient of Differences in Additional values (QDA) 57
4.2.5 The use of the dimension indicators in the present study 58
4.3 Dimension indicators for the method SIFASP–ML 59
4.3.1 The Chi–square statistic 61
4.3.2 The Sequential Chi–square Difference Test (SCDT) 62
4.3.3 Root Mean Square Error of Approximation (RMSEA) 63
4.3.4 Expected Cross–Validation Index (ECVI) 65
4.3.5 Check of fit indices for the model with the correct dimension 67
4.3.6 The use of the fit indices as dimension indicators in the present study 67
4.3.7 Checks on the solutions of LISREL 8 68
4.4 Success Criteria 69
4.4.1 The Recovery Rate (RR) 69
4.4.2 The Difference in Factor correlations (DFC) 70
4.4.3 Additional criteria for the SCA–methods: Amounts of explained variance 70
4.5 Pilot experiment to determine the preferable range of noise to signal ratios for constructed data, using the SCA–methods 71

4.5.1	Data	71
4.5.2	Measures	72
4.5.3	Results	73
4.5.4	Conclusion about meaningful noise levels	75
4.6	Pilot experiment to determine the preferable convergence value for the iterative procedures in the methods SCA-W and SCA-S	75
4.6.1	Data	76
4.6.2	Measures	76
4.6.3	Results	77
4.6.4	Conclusion about meaningful convergence criteria	77

CHAPTER 5 EXPERIMENT 1, FIRST COMPARISON OF THE THREE SCA-METHODS

5.1	Introduction	79
5.2	Manipulations of the data	79
5.2.1	One-population data	80
5.2.2	Two-or-more-populations data	81
5.3	Dependent Variables	81
5.4	Analysis	83
5.5	Questions to be answered by this experiment	85
5.6	Results	86
5.6.1	Finding the preferable dimension indicator	87
5.6.2	Finding the SCA-method with the best dimension indication	91
5.6.3	Finding the SCA-method with the best recovery of factors	96
5.6.3.1	Samples from one population	101
5.6.3.2	Samples from two populations	103
5.6.4	Finding the SCA-method with the best recovery of correlations between factors	108
5.6.5	Discerning the 'samples from one population' from the 'samples from two populations' condition	110
5.6.6	Recovery of factors for specific cases of samples from one population	115

5.7 Conclusion 118

CHAPTER 6 EXPERIMENT 2, TESTING THE METHOD SIFASP-ML

6.1 Introduction 121
6.2 Manipulations of the data 121
6.3 Dependent Variables 122
6.4 Analysis 122
6.5 Questions to be answered by this experiment 124
6.6 Results 124
6.6.1 Finding the preferable dimension indicator for SIFASP-ML 125
6.6.2 Determining the method with the best recovery of factors 132
6.6.2.1 Samples from one population 133
6.6.2.2 Samples from two populations 135
6.6.3 Determining the method with the best recovery of correlations between factors 137
6.6.4 Frequency of bad fit for the solution with the correct dimension 139
6.7 Conclusion 141

CHAPTER 7 EXPERIMENT 3, COMPARISON OF THE SCA-METHODS, WHEN THE STRENGTH OF FACTORS WITH THE SAME INTERPRETATION DIFFERS ACROSS GROUPS

7.1 Introduction 143
7.2 Manipulations of the data 143
7.3 Dependent Variables 145
7.3.1 A problem with the method SCA-S in some of the conditions studied 145
7.3.2 A measure used for inspecting whether the relative strength of the recovered factors resembles the relative strength of the true factors: the Quotient of Component Strengths (QCS) 147

7.4	Analysis	148
7.5	Questions to be answered by this experiment	148
7.6	Results Experiment 3a	149
7.6.1	Finding the SCA-method with the best dimension indication	150
7.6.2	Finding the SCA-method with the best recovery of factors	152
7.6.3	Finding the SCA-method with the best recovery of the relative strength of factors	154
7.6.4	Finding the SCA-method with the best recovery of correlations between factors	157
7.7	Results Experiment 3b	159
7.7.1	Finding the SCA-method with the best dimension indication	159
7.7.2	Finding the SCA-method with the best recovery of factors	159
7.7.3	Finding the SCA-method with the best recovery of the relative strength of factors	160
7.7.4	Finding the SCA-method with the best recovery of correlation between factors	162
7.8	Conclusion	162

CHAPTER 8 EXPERIMENT 4, COMPARISON OF THE SCA-METHODS AND SIFASP-ML-C, FOR VARIOUS CONDITIONS

8.1	Introduction	165
8.2	Manipulations of the data	165
8.3	Dependent Variables	166
8.4	Analysis	167
8.5	Questions to be answered by this experiment	167
8.6	Results Experiment 4	168
8.6.1	Experiment 4a: Four groups	168
8.6.2	Experiment 4b: Samples of n=300, n=500 and n=1000	169
8.6.3	Experiment 4c: Overlapping factors	172
8.7	Conclusion	173

CHAPTER 9 CONCLUSION AND DISCUSSION

9.1	Overview of the results and discussion	175
9.1.1	The preferable method: SCA–P	175
9.1.2	Characteristics of the methods SCA–S, SCA–W and SIFASP–ML	177
9.1.3	Results for the dimension indicators	179
9.1.4	The discriminant function for discerning samples from one population from samples from two populations	181
9.2	Limitations of the present study	182
9.2.1	Beyond the values of the independent variables	182
9.2.2	Data simulation	183
9.2.3	Methods and rotations	184
9.3	User friendliness of the programs used	184
9.4	Guidelines when analyzing data from two or more groups	186

Appendix A	True pattern matrices (of order 12×2 and 12×4) used in this study	189
Appendix B	Identification constraints used for the LISREL program	192
Appendix C	Deriving the largest possible difference in explained variance between the methods SCA–W and SCA–P in case two rank q matrices are analyzed	194
Appendix D	Examining real life data	195
	D.1 Meredith (1964) data	195
	D.2 Van Schuur (1984) data	199

References 205

Summary (in Dutch) 211

Author index 217