Canonical Variate Analysis and Related Methods with Longitudinal Data

Michael Beaghen

Dissertation submitted to the Faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Statistics

Eric P. Smith, Chair
Jesse C. Arnold
Robert V. Foutz
Donald R. Jensen
Keying Ye

November 13, 1997
Blacksburg, Virginia

Keywords: Redundancy Analysis, Procrustes Rotation, Common Principal Components

Canonical Variate Analysis with Longitudinal Data

Michael Beaghen

(ABSTRACT)

Canonical variate analysis (CVA) is a widely used method for analyzing group structure in multivariate data. It is mathematically equivalent to a one-way multivariate analysis of variance and often goes by the name of canonical discriminant analysis. Change over time is a central feature of many phenomena of interest to researchers. This dissertation extends CVA to longitudinal data. It develops models whose purpose is to determine what is changing and what is not changing in the group structure. Three approaches are taken: a maximum likelihood approach, a least squares approach, and a covariance structure analysis approach. All methods have in common that they hypothesize canonical variates which are stable over time.

The maximum likelihood approach models the positions of the group means in the subspace of the canonical variates. It also requires modeling the structure of the within-groups covariance matrix, which is assumed to be constant or proportional over time. In addition to hypothesizing stable variates over time, one can also hypothesize canonical variates that change over time. Hypothesis tests and confidence intervals are developed.

The least squares methods are exploratory. They are based on three-mode PCA methods such as the Tucker2 and parallel factor analysis. Graphical methods are developed to display the relationships between the variables over time.

Stable variates over time imply a particular structure for the between-groups covariance matrix. This structure is modeled using covariance structure analysis, which is available in the SAS package Proc Calis.

Methods related to CVA are also discussed. First, the least squares methods are extended to canonical correlation analysis, redundancy analysis, Procrustes rotation and correspondence analysis with longitudinal data. These least squares methods lend themselves equally well to data from multiple datasets. Lastly, a least squares method for the common principal components model is developed.

Dedicated to my parents.

ACKNOWLEDGEMENTS

Foremost I thank my advisor for his superb guidance. I also thank my committee members for their efforts and the Statistics faculty for their fine teaching and support.

TABLE OF CONTENTS

LIST OF FIGURES x

LIST OF TABLES xi

Notation xii

CHAPTER ONE

INTRODUCTION 1

CHAPTER TWO

BACKGROUND 4

2.1 THE SINGULAR VALUE DECOMPOSITION 5

2.2 CANONICAL CORRELATION AND RELATED MODELS 5
2.2.1 Canonical Correlation 6
2.2.2 Canonical Variate Analysis 7
2.2.3 Correspondence Analysis 9
2.2.4 Redundancy Analysis 10
2.2.5 Procrustes Rotation 11

2.3 THREE-MODE PRINCIPAL COMPONENT ANALYSIS 11
2.3.1 The Tucker3 Model 12
2.3.2 Special Cases of Three-Mode Principal Components 12

2.4 THE CAMPBELL AND TOMENSON MODEL 13

2.5 SUMMARY 14

CHAPTER THREE

PRELIMINARY RESULTS FOR PARAFAC WITH ORTHOGONALITY CONSTRAINTS 15

3.1 INTRODUCTION 15

3.2 OPTIMALITY PROPERTIES OF THE PARAFAC MODEL WITH ORTHOGONALITY CONSTRAINTS 16

3.3 THE NESTEDNESS PROPERTY OF PARAFAC SOLUTIONS WITH ORTHOGONALITY CONSTRAINTS 18

CHAPTER FOUR

COMMON PRINCIPAL COMPONENTS 22

4.1 COMMON PRINCIPAL COMPONENTS 23

4.2 THE LEAST SQUARES APPROACH TO COMMON PRINCIPAL COMPONENTS 25

4.3 LEAST SQUARES APPROACHES TO PARTIAL COMMON PRINCIPAL COMPONENTS AND COMMON SPACE ANALYSIS 30

4.4 COMPARING THE LEAST SQUARES AND MAXIMUM LIKELIHOOD APPROACHES 32

4.5 COMMON COMPONENTS WHICH MAXIMIZE VARIANCE 33

CHAPTER FIVE

RELATING TWO SETS OF VARIABLES OVER A THIRD MODE 36

5.1 INTRODUCTION 36

5.2 RELATING TWO SETS OF VARIABLES OVER A THIRD MODE 37
5.2.1 Redundancy Analysis over a Third Mode 38
5.2.2 Canonical Variate Analysis over a Third Mode 39
5.2.3 Canonical Correlation Analysis over a Third Mode 40
5.2.4 Procrustes Rotation over a Third Mode 41
5.2.5 Which Transformations to Use 42

5.3 HOW TO EVALUATE THE FIT OF THE MODEL 43

5.4 AN EXAMPLE 44

5.5 SOME FURTHER CONSIDERATIONS **52**
5.5.1 Autocorrelation 52
5.5.2 Cross Occasion Covariances 52
5.5.3 Invariance of the Rank of the Solution to Non-Singular Transformations 53
5.5.4 Concluding Comments 53

CHAPTER SIX

GRAPHICAL METHODS 54

6.1 INTRODUCTION **54**

6.2 BIPLOTS **55**
6.2.1 Biplots for Canonical Correlation Analysis 55
6.2.2 Biplots for Redundancy Analysis 57
6.2.3 Biplots for Procrustes Rotation 58

6.3 JOINT PLOTS **59**
6.3.1 Joint Plots for Canonical Correlation Analysis 59
6.3.2 Joint Plots for Redundancy Analysis 61
6.3.3 Joint Plots for Procrustes Rotation 64

6.4 PLOTS OF THE COMPONENT SCORES **65**

6.5 RESIDUAL PLOTS **71**

6.6 SUMMARY **72**

CHAPTER SEVEN

COVARIANCE STRUCTURE ANALYSIS 74

7.1 INTRODUCTION **74**

7.2 COVARIANCE STRUCTURE ANALYSIS **75**

7.3 MODELING CANONICAL VARIATE ANALYSIS OVER TIME AS A COVARIANCE STRUCTURE **76**

7.4 PUTTING CVA OVER TIME IN THE COSAN FRAMEWORK **77**

7.5 AN EXAMPLE **78**

7.6 CONCLUDING REMARKS **80**

CHAPTER EIGHT

CANONICAL VARIATE ANALYSIS OVER TIME 81

8.1 INTRODUCTION 81

8.2 PRELIMINARIES 82
8.2.1 Orthogonal Versus Uncorrelated Variates 82
8.2.2 The Structure of the Data 83

8.3 THE CVA/TIME (ORTHOGONAL) MODEL 83
8.3.1 The CVA/Time Model with Orthogonal Variates 85
8.3.2 Sufficient Statistics 86
8.3.3 Estimating Equations 87
8.3.4 Unchanging Group Positions 91
8.3.5 Obtaining Estimates 91
8.3.6 Statistical Inference 91

8.4 SIMULATIONS 93

8.5 CVA/TIME - UNCORRELATED VARIATES 97
8.5.1 The CVA/Time Model with Uncorrelated Variates 98
8.5.2 Estimating the Within-Groups Covariance Matrix 99
8.5.3 Estimating the Matrix of Proportionality Constants (A) 101
8.5.4 Hypothesis Test for the Simple Structure of the Covariance Matrix 102
8.5.5 Estimating the Canonical Variates and the Group Scores 103
8.5.6 Estimating Unchanging Group Positions 105

8.6 EXAMPLE FOR CVA/TIME WITH UNCORRELATED VARIATES - SEX DIFFERENCES IN MATH ANXIETY BEFORE AND AFTER INTRODUCTORY CALCULUS 106

8.7 A COMPARISON TO ALTERNATIVE METHODS, INCLUDING DOUBLY MULTIVARIATE REPEATED MEASURES 113
8.7.1 Two Simple Approaches 114
8.7.2 Measurements at Different Occasions Treated as Distinct Variables 114
8.7.3 Doubly Multivariate Repeated Measures 116

CHAPTER NINE

SCALING THE VARIABLES 119

9.1 AN EXAMPLE OF RESCALING THE DATA 119

9.2 DEFINITIONS OF SCALE INVARIANCE 120

9.3 EXAMPLES OF SCALE INVARIANT METHODS 121

9.4 SCALE INVARIANCE FOR THREE-MODE PRINCIPAL COMPONENTS ANALYSIS 123

9.5 HOW TO SCALE THE DATA 125

CHAPTER TEN

CONCLUSION AND FURTHER RESEARCH 127

BIBLIOGRAPHY 130

APPENDIX ONE: THE KRONECKER PRODUCT OF TWO MATRICES 134

APPENDIX TWO: THE F-G ALGORITHM 135

APPENDIX THREE: THE PROGRAMS FOR PARAFAC (ORTH.) AND TUCKER2 FOR THE SHENANDOAH EXAMPLE 137

APPENDIX FOUR: CODE FOR PLOTS AND GRAPHS 141

APPENDIX FIVE: THE PROGRAM CODE FOR THE COSAN MODEL 144

APPENDIX SIX: SAS CODE FOR THE SIMULATION OF THE COMMON VARIATES MODEL 154

APPENDIX SEVEN: MATHEMATICA CODE FOR CALCULATING THE ASYMPTOTIC COVARIANCE MATRIX OF THE ESTIMATES 161

APPENDIX EIGHT: ASYMPTOTIC COVARIANCE MATRIX FOR THE PARAMETER ESTIMATES 163

APPENDIX NINE: COVARIANCE MATRIX OF PARAMETER ESTIMATES BASED ON THE SIMULATION 164

APPENDIX TEN: THE COVARIANCE MATRIX FOR THE ITEMS ON THE MATH ANXIETY QUESTIONNAIRE 165

VITA 167

LIST OF FIGURES

Figure 2.1 Scatterplots in the Untransformed Space 9
Figure 2.2 Scatterplots in the Transformed Space 9
Figure 6.1 Joint Plot for the Sum of Core Matrices for PARAFAC (orth.) 63
Figure 6.2 Key to Symbols 63
Figure 6.3 Scores on the First Streamwater Variate 67
Figure 6.4 Scores on the Second Streamwater Variate 68
Figure 6.5 Scores on the Third Streamwater Variate 69
Figure 6.6 Scores on the Fourth Streamwater Variate 70
Figure 6.7 Residual Plot for the Sums of Squares Explainable of the Streamwater Variables 72
Figure 6.8 Key to Symbols 72
Figure 7.1 Estimates for the COSAN and PARAFAC Models 79
Figure 8.1 Group Data in Untransformed Space - First Occasion 84
Figure 8.2 Group Data in Untransformed Space - Second Occasion 85
Figure 8.3 Group Data in transformed Space - First Occasion 98
Figure 8.4 Group Data in transformed Space - Second Occasion 98
Figure 8.5 The Positions of the Group Means 109
Figure 8.6 Men's Scores at the First Occasion 110
Figure 8.7 Women's Scores at the First Occasion 110
Figure 8.8 Men's Scores at the Second Occasion 111
Figure 8.9 Women's Scores at the Second Occasion 111
Figure 8.10 The Matrix of Proportionality Constants (**A**) 113

LIST OF TABLES

Table 5.1 Error Variance of Responses 46
Table 5.2 PARAFAC (orth.) Core 47
Table 5.3 Core Matrices for the Tucker2 47
Table 5.4 Canonical Variate X-weights for PARAFAC (orth.) 48
Table 5.5 Canonical Variate X-weights for the Tucker2 48
Table 5.6 Y-weights for PARAFAC 49
Table 5.7 Y-weights for Tucker2 49
Table 5.8 Matrix of Sums of Squares Explained by Variable and Component 50
Table 8.1 Parameter Estimates 95
Table 8.2 Theoretical and Observed Variances for the Parameter Estimates 96
Table 8.3 Theoretical and Observed Values of the Likelihood Ratio Test Statistic 97
Table 8.4 The Math Anxiety Questions 106
Table 8.5 Possible Responses 106
Table 8.6 Canonical Variate Weights and Structural Coefficients 108
Table 8.7 Observed and Predicted Means at the First Occasion 112
Table 8.8 Group Means and Standard Deviations for each Question 113

Notation

There are several conventions that are adhered to in this dissertation. A column vector is bold and small lettered. Matrices are bold and large lettered. Scalars are small lettered and not bold. A constant is italicized, an index is not.