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
Copyright 1997, Michael Beaghen
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
CHAPTER EIGHT

CANONICAL VARIATE ANALYSIS OVER TIME

8.1 INTRODUCTION

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

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

8.4 SIMULATIONS

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

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

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

CHAPTER NINE

SCALING THE VARIABLES

9.1 AN EXAMPLE OF RESCALING THE DATA

9.2 DEFINITIONS OF SCALE INVARIANCE

9.3 EXAMPLES OF SCALE INVARIANT METHODS

9.4 SCALE INVARIANCE FOR THREE-MODE PRINCIPAL COMPONENTS ANALYSIS

9.5 HOW TO SCALE THE DATA
# LIST OF FIGURES

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