CHAPTER ONE

INTRODUCTION

Change over time is a central feature of many phenomena of interest to researchers. Analysis of data measured over time is well developed for univariate analysis (Hand & Crowder 1989, Diggle, Liang & Zeger 1994), but not as well developed for multivariate analysis. In this dissertation I extend certain multivariate methods to longitudinal data so that one can investigate the change over time in the underlying structures in the data.

The main interest of the dissertation is extending canonical variate analysis (CVA) to longitudinal data. However, other related multivariate methods which I extend to longitudinal data are canonical correlation analysis (CCA), redundancy analysis (RA) and Procrustes rotation (PR). These are kindred methods which relate two sets of variables, which I shall designate X-variables and the Y-variables. In the case of CVA the X-variables are group indicators, while for CCA, RA and PR the variables may be either continuous or categorical. Of these four methods, I give particular attention to canonical variate analysis because of its obvious usefulness and because it is the most mathematically tractable when extended to data taken over time.

In addition to change over time I investigate change that occurs over different datasets or groups. Although not a primary interest of this dissertation, the problem of modeling canonical correlation analysis, canonical variate analysis, redundancy analysis and Procrustes rotation with data from multiple datasets is closely related to the problem of modeling these analyses with longitudinal data. I also devote a chapter to common principal components analysis (Flury 1988), which models principal components with data from multiple datasets and shares conceptual similarities to the other models I discuss.

Several multivariate models for data over time have already been developed. Related to some of the approaches to be taken in this dissertation, Kiers (1991) gives an overview of using three-mode principal components to model principal components over time. Three-mode principal components decomposes three-mode data, that is, data which is in the form of a three-way array. Swaminathan (1984) has developed models for factor analysis with longitudinal data. Modeling the relationship between two sets of variables over time, however, has not been well developed. Regression can be viewed as a model relating two sets of variables, where one set consists of the response. Regression over time has been modeled by different methods. Ware (1985) describes random effects models, autoregressive models and multivariate models. Liang and Zeger (1986) discuss using generalized linear models for longitudinal regression. Akin to factor analysis and multiple regression is LISREL (LInear Structural RELations, Jöreskog 1989), which relates variables in a structural model. Jöreskog (1979) describes how LISREL can be approached with longitudinal data, although the results are limited.

In addition to multivariate models for longitudinal data, a few multivariate methods for data from multiple datasets or groups have also been developed. A method for performing principal components analysis on data from multiple datasets is common principal components (see Chapter Four). Closely related to one method developed in this dissertation is a method for performing canonical variate analysis on data from multiple datasets (Campbell and Tomenson 1983).

What is shared by the models mentioned above is that they hypothesize common or stable variates across the multiple occasions or datasets. I shall take the same approach to modeling change. The common component or variate shall be the *leitmotiv* of this dissertation. The nature of change shall be investigated by asking the question: What remains stable over time, and what changes? In particular, I ask if it is useful to model variates as constant or stable over time. If the common variate approach is deemed useful, then the nature of the change is indicated by the strength of the relationship, i.e., the orrelation or covariance, between the pairs of variates at each occasion.

Having outlined the problem and the nature of the attempted solution, I will now lay out the organization of the dissertation. Chapter Two provides the reader with the background material necessary to frame the problem and approach the solutions attempted in the rest of the dissertation. Chapter Two first discusses canonical correlation analysis, followed by canonical variate analysis, which is a special case of canonical correlation analysis. Then it discusses redundancy analysis and Procrustes rotation, two methods closely related to canonical correlation analysis. Next three-mode methods are discussed. Lastly, Campbell and Tomenson's model for canonical variate analysis for multiple datasets is discussed.

While the material presented in Chapter Two is strictly a review, the chapters which follow present mostly new material. Background material appears in Chapters Four, Six, Seven and Nine. These chapters are self-contained; they have both new material and the related background.

In Chapter Three I show the partitioning of sums of squares and prove the nestedness of the solutions for the PARAFAC (orth.) model. These results are important for the developments of Chapters Four and Five.

Chapter Four compares the maximum likelihood and least squares approaches to common principal components (CPC). It starts with background on Flury's (1984) maximum likelihood approach, then outlines how CPC can be approached by three-mode PCA. This chapter is the one least organically connected to the rest of the dissertation. Besides interest in the CPC model for its own sake, it relates to the main thesis in two ways. First, it is a clear exposition of a common variate model. Second, it shows how maximum likelihood and least squares methods complement each other. In this sense it suggests possible developments of common variate models for canonical variate analysis, canonical correlation analysis and redundancy analysis over time which could be approached by both maximum likelihood and least squares.

Chapters Five and Six present least squares methods. Chapter Five models canonical variate analysis, canonical correlation analysis, redundancy analysis and Procrustes rotation over multiple occasions and over multiple datasets with three-mode principal components. In Chapter Six graphical techniques are developed to be used in conjunction with the least squares methods of Chapter Five. I do not develop hypothesis tests for the least squares methods. Hence the least squares methods are not as powerful as the maximum likelihood method. However, they are more flexible. They can be applied to data where the X-variables are continuous. They can be applied to data where Y-variables are categorical, such as correspondence analysis. Furthermore, they are well suited for exploratory analysis.

In Chapter Seven canonical variate analysis over time is approached by the analysis of covariance structures or COSAN (COvariance Structure ANalysis, McDonald 1978), which is related to LISREL modeling.

In Chapter Eight I develop a model for canonical variates over time which is estimated by maximum likelihood. Like Campbell and Tomenson's approach (see Section **2.5**), it models group means. This is the only maximum likelihood method that I develop fully in the dissertation. I develop hypothesis tests based on the likelihood ratio principal and confidence intervals based on the inverse of the information matrix.

In Chapter Nine I discuss the important issue of the scaling of the variables. Chapter Ten concludes the dissertation and suggests further research.