

Overview of Multivariate Methods

James L. Schmidhammer

Multivariate Analysis

- Refers to the analysis of multiple response variables
- Concerned with finding relationships among
 - The response variables
 - The experimental units
 - both

Multivariable Techniques

- Refer to situations where each experimental unit is measured on multiple explanatory variables, but only one response variable

Relationships among Response Variables

Example:

Subtest scores on intelligence or personality tests

Relationships among Experimental Units

Example:

Nutritional content of cereals measured on
fat, protein, carbohydrates, and sodium

Multivariate Techniques

- Tend to be exploratory
- Assume independence of experimental units

Principal Components Analysis

- Transforms a set of correlated response variables into a new set of uncorrelated variables
- Used for the reduction of dimensionality (by using only the first several principal components)
- The principal components themselves are generally uninterpretable

Principal Components Analysis

A dimension reduction technique

- creates new variables that are linear combinations of a set of correlated variables
- does not assume an underlying latent factor structure.

Practical question:

“How can I reduce the set of 20 correlated variables to a more manageable number of uncorrelated variables?”

Chapter 5

Principal Components Analysis

- Useful as a tool for screening multivariate data
- Useful as an intermediate step before
 - Cluster analysis
 - Multivariate regression
 - Factor analysis

PCA Example: Bird Habitat

A researcher is interested in understanding the habitat of a species of bird. Twenty characteristics are measured for each habitat. Many of these measures are associated.

These variables will be used in regression, discriminant, and cluster analyses.

The researcher wants to reduce the total number of variables from 20 to something smaller and eliminate potential collinearity problems.

Factor Analysis

- Used to create a (smaller) set of new variables that contains (nearly) all of the information in an original set of correlated response variables
- Initial development with IQ tests
 - Pointed to an overall measure of intelligence
- Intent is to create factors (i.e., the new variables) that are interpretable.

Factor Analysis

Exploratory factor analysis is a variable identification technique with superficial resemblance to principal components analysis, but with some important distinctions.

Factor analytic methods are used when an underlying factor structure is presumed to exist but cannot be represented easily with a single value.

Chapter 5

Factor Analysis Example: Perceptions of Mathematics in School

An researcher wants to know whether students' self-perceptions in math reflect several underlying latent factors or one single factor.

Questionnaire items from several instruments intended to measure mathematics-related perceptions are administered to 4000 students.

- Exploratory analysis identifies possible underlying factors.
- Confirmatory analysis tests hypotheses about factors.

Correspondence Analysis

- Roughly, a technique conceptually similar to Factor Analysis, but used on non-metric variables (i.e., variables measured on a nominal or ordinal scale)

Structural Equation Modeling

Also called Confirmatory Factor Analysis

SEM is used to investigate regression-type relationships among a set of observed or latent responses and a set of observed or latent predictors.

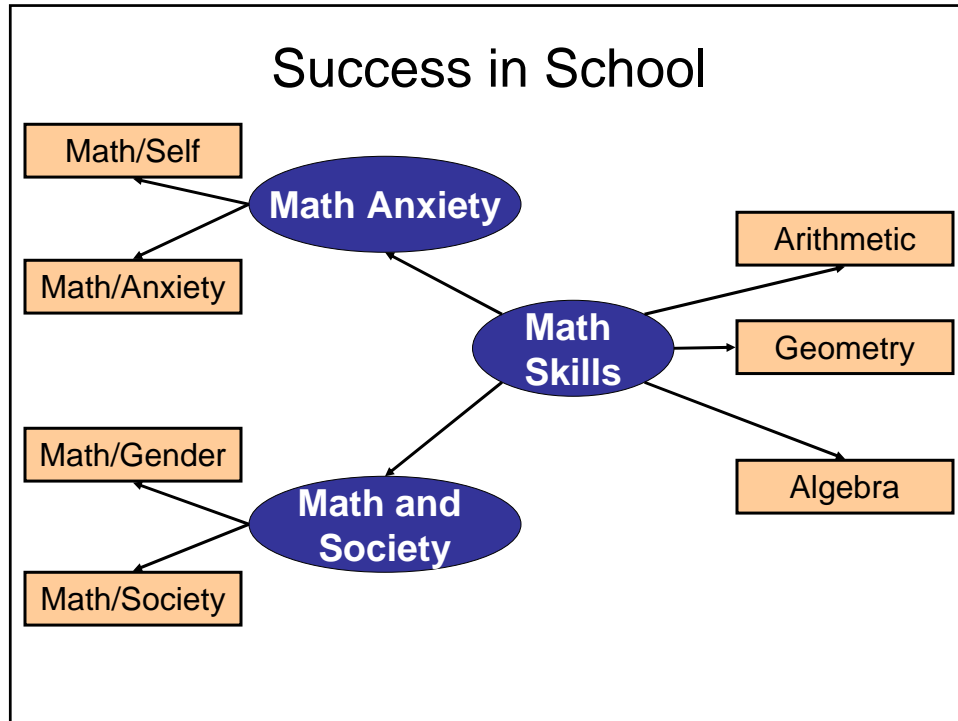
Multivariate multiple regression, path analysis, and confirmatory factor analysis are all special cases of structural equation models.

Chapter 6

Structural Equation Modelling Example: Success in Math

A research team wants to know whether math anxiety and students' perceptions of math in society are associated with mathematical ability among high school students.

Math anxiety, perceptions of math in society, and mathematical ability are all **latent variables**: they cannot be measured directly but must be estimated with a combination of measured variables.



Discriminant Analysis

- Multiple response variables measured on experimental units in 2 (or more) predefined groups.
- Determines how to use these variables to best discriminate between the groups.
- Also used to classify new observations into one of the two groups.
- Developed by C. R. Rao. Early applications include successful discrimination between the Aryan and Dravidian populations in India.

Discriminant Analysis

Discriminant analysis is a dimension reduction method that can be used to identify a linear combination of variables that produces the greatest distance between categories. *Discriminant analysis* is conceptually similar to logistic regression for multivariate data, and it is computationally similar to MANOVA.

Chapter 4

Multiple Discriminant Analysis

- Discriminant Analysis involving more than two groups

Canonical Discriminant Analysis

- Principal components used as variables for discrimination

Discriminant Analysis Example: Pathological Gambling

Researchers want to use responses to questionnaire items to classify people identified as steady gamblers, binge gamblers, and control/non-gamblers.

A twelve-item questionnaire is administered to three groups of participants.

Question: What linear combination of responses accounts for most of the variation in classification of gamblers?

Discriminant Analysis Example: Customer Profiling and Prediction

A credit card company is interested in using financial information to decide whether potential customers represent good or bad risk before offering a credit card.

An analyst is interested in understanding what combination of demographic variables best predict whether a customer prefers one of several different marketing strategies.

Logistic Regression

- Really a multivariable technique
 - One dichotomous dependent variable
 - Multiple explanatory variables
- Used as an alternative to Discriminant Analysis in order to classify observations into groups
- Less assumptions than Discriminant Analysis

Cluster Analysis

- Used to develop meaningful subgroups (or clusters) of experimental units
- Initial number of clusters is unknown
- Can also be used to cluster variables
 - An alternative to factor analysis

Multidimensional Scaling

- Maps distances between points in a high dimensional space into a lower dimensional space

Multivariate Analysis of Variance (MANOVA)

- Simultaneous ANOVA with multiple response variables
- Controls overall Type I Error Rate

Multivariate Analysis of Variance (MANOVA)

Multivariate analogy to ANOVA.

- Tests for significant differences between groups on two or more related dependent variables simultaneously while accounting for the correlation among the dependent variables.
- Research question: “Are there significant differences between two or more groups on a set of responses?”

Chapter 2

MANOVA Example: A Comparison of Advertising Strategies

Evaluate the effectiveness of three different commercials.

What makes an effective commercial?

- Product recognition
- Product recall
- Product liking
- Price willing to pay for product

This example has

- three levels of the independent variable
- four response variables.

MANOVA Example: The Effectiveness of a Drug

A drug company wants to compare the effectiveness of two different drug formulations (Old, New) across different dosages (50, 100, 200 mg).

How is effectiveness evaluated?

- Score on a depression scale
- Scores on two different obsessive-compulsive behavior scales.

2 × 3 factorial design, three responses.

Canonical Variates Analysis

- MANOVA on Principal Components

Multivariate Regression

- Multiple Regression with multiple dependent variables

Multivariate Multiple Regression

Test for significant linear relationship between a set of predictors and a set of responses while accounting for the correlations among the responses.

Research Question:

“Does variation in a set of continuous independent variables adequately predict a set of continuous responses?”

Chapter 3

Multivariate Regression Example: The Diagnostic Usefulness of an Instrument

How well does a new psychological instrument perform compared to an established instrument?

- The established instrument, based on diagnostic criteria, contains twelve items and must be administered by a trained interviewer.
- The test instrument contains twenty items and can be completed by the respondent with pen and paper.

This example has

- twelve continuous predictors
- twenty continuous responses.

Canonical Correlation Analysis

- Measures the correlation between one set of variables and a different set of variables
- Multivariate generalization of the simple Pearson Product Moment Correlation

Canonical Correlation Analysis

Canonical correlation analysis tests the same hypotheses as multivariate regression, but also allows you to

- interpret how the predictors are related to the responses
- interpret how the responses are related to the predictors
- examine how many dimensions the variable sets share in common.

Chapter 3