

DSC 410/510 Multivariate Statistical Methods

Introduction

1

What is DSC 410/510?

- Applications-oriented introduction to multivariate statistical methods for MBAs and upper-level business undergraduates
- Focus: design, estimation, interpretation
- Four techniques considered in depth:
 - Conjoint analysis
 - Cluster analysis
 - Discriminant analysis / logistic regression
 - Multidimensional scaling / correspondence analysis

2

Computing

- Current computer technology enables many multivariate problems to be analyzed
- This class: SAS statistical software
- Alternatives: SPSS, Conjoint Analyzer
- HATCO dataset
- Specialized conjoint data
- Your own data for class-work & projects!

3

What is Multivariate Analysis?

- Statistical methods that simultaneously analyze multiple variables on each case
- Variables are inter-related such that their effects cannot be interpreted separately
- Most business areas include multivariate techniques in their analytical toolbox
 - e.g. marketing: conjoint analysis

4

Some Quotes

- [Multivariate analysis methods] make it possible to ask specific and precise questions of considerable complexity in natural settings. ... The natural correlations among the manifold influences on behavior can be preserved and separate effects of these influences can be studied statistically.
- [Businesspeople in most markets today] must develop strategies to appeal to numerous segments of customers with varied demographic and psychographic characteristics in a marketplace with multiple constraints (e.g. legal, economic, competitive, technological).

5

The Variate

- A linear combination of variables with empirically determined weights:
$$w_1X_1 + w_2X_2 + w_3X_3 + \dots + w_pX_p$$
- For example, in discriminant analysis, weights are determined to maximally differentiate among groups of cases
- Also of interest are each separate variable's contribution to overall variate effect

6

Measurement Scales

- Metric (i.e. Quantitative)
 - ◆ Values indicate an amount or quantity of the variable for each case
 - ◆ e.g. temperature or weight
- Nonmetric (i.e. Qualitative)
 - ◆ Values indicate which of a fixed number of categories each case belongs to
 - ◆ e.g. gender or consumer satisfaction

7

Metric Measurement Scales

- Interval
 - ◆ arbitrary zero point
 - ◆ e.g. temperature in Fahrenheit
- Ratio
 - ◆ absolute zero point
 - ◆ e.g. weight in pounds

8

Nonmetric Measurement Scales

- Ordinal
 - ◆ categories have a meaningful order
 - ◆ e.g. consumer satisfaction on a 5-point scale
- Nominal
 - ◆ categories have no meaningful order
 - ◆ e.g. gender or political party

9

Why Worry?

- Analyses designed for metric scales often produce gibberish with nonmetric scales, and vice versa
- The most appropriate technique depends on the measurement scale for each variable

10

Representing Nonmetric Data

- Analyzing nonmetric data requires values to be represented by numbers
- Nominal scales use dummy/indicator variables
 - ◆ e.g. $D_1 = 0$ if male, 1 if female
 - ◆ e.g. $D_2 = 1$ if democrat, 0 otherwise
 $D_3 = 1$ if republican, 0 otherwise
- Ordinal scales
 - ◆ either use dummy variables, but often with adapted techniques
 - ◆ or are analyzed as metric data, if justified

11

Measurement Error

- Difference between what is measured and the “true value”
- For example: data entry errors, too refined rating scales, inaccurate survey responses
- Measurements should be **valid** to accurately represent the concept in question
- Measurements should also be **reliable** (i.e. consistent), and so minimize error

12

Summated Scales

- Several variables or indicators joined in a composite measure to represent a concept
- For example, summated ratings of product satisfaction: features, color, value, etc.
- Published marketing scales are available that have demonstrated reliability:
 - ◆ Bearden, Netemeyer and Mobley (1993)
 - ◆ Brunner and Hensel (1993)

13

Significance and Power

- Type I error: reject H_0 when H_0 true
 - ◆ Set $\alpha = \text{Pr}(\text{Type I error})$
- Type II error: not reject H_0 when H_0 false
 - ◆ Power = $1 - \beta = 1 - \text{Pr}(\text{Type II error})$
- Power depends on:
 - ◆ effect size
 - ◆ α
 - ◆ sample size

14

Techniques Covered in DSC 410/510

- Conjoint analysis
- Cluster analysis
- Discriminant analysis
 - ◆ Logistic regression
- Multidimensional scaling
 - ◆ Correspondence analysis

15

Examples

Objectives	Design
1. Segment customers into groups with similar perceptions of Hatco	Customers rated performance of Hatco on 7 attributes
2. Develop new cleanser and understand customer preferences	Customers rated 22 cleansers with different attributes
3. Explore perceptions of Hatco relative to 9 major competitors and how comparisons are made	Customers rated similarities of Hatco and each competitor, and also 8 attributes with each firm
4. Identify perceptions of Hatco that differ between 2 types of firm	Customers for each firm type rated performance of Hatco on 7 attributes

16

Conjoint analysis

- Used to evaluate new products, services, or ideas, and to develop new products/services
- Consumers evaluate a few product profiles (combinations of attribute levels)
- Researcher can then assess importance of attributes, as well as levels of attributes
- Results can also be used in design simulators to help find optimal products

17

Cluster Analysis

- Used to develop clusters (groups) of cases, e.g. individuals, products
- Clusters are *not* predetermined
- Cases within clusters are *similar* somehow
- Three-step analysis:
 1. Determine measure of similarity
 2. Partition cases into clusters
 3. Profile cases within clusters

18

Multiple Discriminant Analysis

- Used to characterize different groups in a population
- Group membership identified by a nonmetric dependent variable
- Characterization based on metric independent variables
- Results enable understanding of group differences and prediction of group membership

19

Logistic Regression

- Similar to multiple linear regression in that independent (predictor) variables are used to predict a dependent (response) variable
- However, here response is nonmetric (in particular either binary or binomial)
- Predictor variables can be both metric and nonmetric
- Multivariate normality not required (unlike discriminant analysis)

20

Multidimensional Scaling

- Used to represent consumer preferences or similarity scores on a perceptual map
- For example, consider similarities of six suppliers (labeled A-F)
 - ◆ if A and B considered most similar, they will be closest together in resulting map
 - ◆ results can clarify underlying relationships

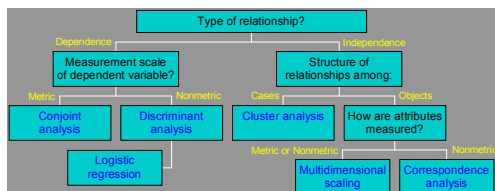
21

Correspondence Analysis

- Used to display the relationships in a cross-tabulation of two categorical variables
- For example, consider sales for 4 products broken down by 3 age categories
 - ◆ perceptual map shows similarities of different products and age categories
 - ◆ map also shows associations of particular products with each age category

22

Classification of Techniques



A dependence relationship has a (dependent) variable to be predicted or explained by other (independent) variables

23

Examples Revisited

Objectives

1. Segment customers into groups with similar perceptions of Hatco
2. Develop new cleanser and understand customer preferences
3. Explore perceptions of Hatco relative to 9 major competitors and how comparisons are made
4. Identify perceptions of Hatco that differ between 2 types of firm

Design

Customers rated performance of Hatco on 7 attributes

Customers rated 22 cleansers with different attributes

Customers rated similarities of Hatco and each competitor, and also 8 attributes on each firm

Customers for each firm type rated performance of Hatco on 7 attributes

24

Multivariate Analysis Guidelines

- Significance: practical (substantive effects that justify action) as well as statistical
- Samples can be too small (low power, poor generalizability) or large (overly sensitive tests)
- Examine your data: plots, diagnostic techniques, assess assumptions, look at prediction errors
- Avoid unnecessarily complicated models or inclusion of irrelevant variables
- Validate model to ensure generalizability

25

Decision Framework

1. Define problem/objectives and select technique
2. Develop research design specific to problem
3. Evaluate assumptions (statistical and conceptual)
4. Estimate model and assess the fit
5. Interpret results (e.g. consider the variate)
6. Validate model to assess generalizability

Usually iterate stages 4 and 5

Stage 3 can often be done only *after* estimating model

26