

海外优秀数学类教材系列丛书

THOMSON

影印版

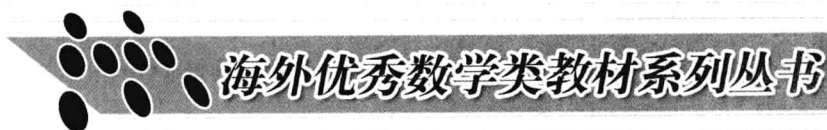
*Applied Multivariate Methods  
for Data Analysts*

# 应用多元统计 分析方法

□ DALLAS E. JOHNSON



高等教育出版社  
Higher Education Press



影印版

# *Applied Multivariate Methods for Data Analysts*

# 应用多元统计分析方法

Dallas E. Johnson

*Kansas State University*



高等教育出版社  
Higher Education Press

图字: 01-2004-3218 号

Dallas E.Johnson

Applied Multivariate Methods for Data Analysts, first Edition

ISBN: 0-534-23796-7

Copyright © 1998 by Duxbury, a division of Thomson Learning

Original language published by Thomson Learning (a division of Thomson Learning Asia Pte Ltd). All Rights reserved. 本书原版由汤姆森学习出版集团出版。版权所有, 盗印必究。

Higher Education Press is authorized by Thomson Learning to publish and distribute exclusively this English language reprint edition. This edition is authorized for sale in the People's Republic of China only (excluding Hong Kong, Macao SAR and Taiwan).

Unauthorized export of this edition is a violation of the Copyright Act. No part of this publication may be reproduced or distributed by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

本书英文影印版由汤姆森学习出版集团授权高等教育出版社独家出版发行。此版本仅限在中华人民共和国境内(但不允许在中国香港、澳门特别行政区及中国台湾地区)销售。未经授权的本书出口将被视为违反版权法的行为。未经出版者预先书面许可, 不得以任何方式复制或发行本书的任何部分。

981-265-506-9

#### 图书在版编目(CIP)数据

应用多元统计分析方法 = Applied Multivariate Methods for Data Analysts / (美) 约翰逊 (Johnson, D.

E.) 著. —北京: 高等教育出版社, 2005. 6

(海外优秀数学类教材系列丛书)

ISBN 7-04-016545-7

I. 应... II. 约... III. 多元分析 — 统计分析 — 分析方法 — 高等学校 — 教材 — 英文 IV. 0212.4

中国版本图书馆 CIP 数据核字 (2005) 第 044048 号

策划编辑 徐可 责任编辑 徐可 封面设计 王凌波 责任印制 陈伟光

出版发行 高等教育出版社  
社 址 北京市西城区德外大街 4 号  
邮政编码 100011  
总 机 010-58581000  
  
经 销 北京蓝色畅想图书发行有限公司  
印 刷 北京民族印刷厂  
  
开 本 787 × 1092 1/16  
印 张 36.5  
字 数 650 000

购书热线 010-58581118  
免费咨询 800-810-0598  
网 址 <http://www.hep.edu.cn>  
<http://www.hep.com.cn>  
网上订购 <http://www.landaco.com>  
<http://www.landaco.com.cn>  
  
版 次 2005 年 6 月第 1 版  
印 次 2005 年 6 月第 1 次印刷  
定 价 43.30 元 (含光盘)

本书如有缺页、倒页、脱页等质量问题, 请到所购图书销售部门联系调换。

版权所有 侵权必究

# Preface

I once attended a conference at which George Box stated that “Statistics is much too important to be left entirely to statisticians.” A bit later, Walt Federer stated that “Science is much too important to be left entirely to scientists.” Both of these famous statisticians were correct! Never before in the history of science and statistics has there been a greater need for interactions and collaborations between scientists and statisticians. This book helps to facilitate such collaborations and interactions. I have been fortunate in that I have had substantial contact with scientists during my tenure at Kansas State University. These collaborations have greatly influenced my approach to teaching multivariate methods. I believe that multivariate methods are too important to be taught only to statisticians.

Furthermore, I have been teaching public seminars and college courses in applied multivariate analyses for the last 20 years. In these seminars and courses, students have posed many important questions that multivariate methods can help answer. As data sets grow in size, multivariate methods become ever more useful. Today’s technologies make it very easy to collect large amounts of data; multivariate methods are needed to determine whether such massive amounts of data actually contain information. It has been said that while it is easy to collect data, it is much harder to collect information. Multivariate methods can help determine whether there is information in data, and they can also help to summarize that information when it exists.

To date, textbooks have emphasized only the theory of multivariate methods or only the application of the methods. Readers were given information that was either too advanced to apply or too elementary to illustrate the power of the methods. This text has broken the mold by focusing on the why, when, how, and what of multivariate analyses and answering the following questions:

Why should multivariate methods be used?

When should they be used?

How can they be used?

And what has been learned by the application of the methods?

Ideally, users of this book will have had a previous course in statistics that included multiple regression. Some familiarity with matrix algebra is desirable, but not crucial. My approach assumes familiarity with most of the statistical

concepts encountered in a first course in statistics, such as means and standard deviations, correlations, p-values, hypothesis tests, and confidence limits.

While this text is loaded with examples using real data, several of the exercises are directed at data sets that students are asked to provide from their own experiences. I find that students enjoy working with their own data. So, when I teach multivariate methods, I require each student to provide a data set for class use along with a description of the data's important features and the reasons behind its being collected. These data sets are then placed in a computer directory that every student in the class can access. I then use these data sets as much as possible when assigning exercises to the class. I strongly encourage instructors who use this book to do the same.

Other unique features of this text include:

- annotated computer output, emphasizing SAS and SPSS
- extensive use of graphics to explain concepts
- data disk that contains data files from text discussion and exercises as well as computer commands used to create the analyses described throughout the text

I owe much of the development of this text to those who have participated in my seminars and courses. From these "students," I learned about the needs, their concerns, and their abilities. In writing this text, I have tried to address their needs and concerns, while recognizing their differing abilities.

## Acknowledgments

I wish to express my appreciation to all who helped me with the development of this text. I am particularly grateful to the students at Kansas State University and students who have taken public seminars through the Institute of Professional Education. These students have provided numerous valuable suggestions that have greatly improved the content of this text. I would also like to thank Ms. Carolyn Crockett and Mr. Alexander Kugushev for their valuable suggestions. I would like to thank the following reviewers for their helpful comments: Marcia Gumpertz, North Carolina State University; John E. Hewitt, University of Missouri, Columbia; Linda S. Hynan, Baylor University; Dipak Jain, Northwestern University; Lincoln Moses, Stanford University; Mack C. Shelley II, Iowa State University; Eric Smith, Virginia Polytechnic Institute; and Richard Sundheim, St. Cloud State University. I also thank Mr. Jane Cox for her help in creating many of the formulas in this text. Finally, I would like to thank my parents, Chet and Dorothy Johnson, for giving me the opportunity for furthering my education and my wife, Erma, for the help and support that she provided during this endeavor.

*Dallas Johnson*

# Contents

<b>1.</b>	<b>APPLIED MULTIVARIATE METHODS</b>	<b>1</b>
<b>1.1</b>	<b>An Overview of Multivariate Methods</b>	<b>1</b>
	Variable- and Individual-Directed Techniques	2
	Creating New Variables	2
	Principal Components Analysis	3
	Factor Analysis	3
	Discriminant Analysis	4
	Canonical Discriminant Analysis	5
	Logistic Regression	5
	Cluster Analysis	5
	Multivariate Analysis of Variance	6
	Canonical Variates Analysis	7
	Canonical Correlation Analysis	7
	Where to Find the Preceding Topics	7
<b>1.2</b>	<b>Two Examples</b>	<b>8</b>
	Independence of Experimental Units	11
<b>1.3</b>	<b>Types of Variables</b>	<b>11</b>
<b>1.4</b>	<b>Data Matrices and Vectors</b>	<b>12</b>
	Variable Notation	13
	Data Matrix	13
	Data Vectors	13
	Data Subscripts	14
<b>1.5</b>	<b>The Multivariate Normal Distribution</b>	<b>15</b>
	Some Definitions	15
	Summarizing Multivariate Distributions	16
	Mean Vectors and Variance–Covariance Matrices	16
	Correlations and Correlation Matrices	17
	The Multivariate Normal Probability Density Function	19
	Bivariate Normal Distributions	19

<b>1.6</b>	<b>Statistical Computing</b>	<b>22</b>
	Cautions About Computer Usage	22
	Missing Values	22
	Replacing Missing Values by Zeros	23
	Replacing Missing Values by Averages	23
	Removing Rows of the Data Matrix	23
	Sampling Strategies	24
	Data Entry Errors and Data Verification	24
<b>1.7</b>	<b>Multivariate Outliers</b>	<b>25</b>
	Locating Outliers	25
	Dealing with Outliers	25
	Outliers May Be Influential	26
<b>1.8</b>	<b>Multivariate Summary Statistics</b>	<b>26</b>
<b>1.9</b>	<b>Standardized Data and/or Z Scores</b>	<b>27</b>
	Exercises	28

## 2. SAMPLE CORRELATIONS

<b>2.1</b>	<b>Statistical Tests and Confidence Intervals</b>	<b>35</b>
	Are the Correlations Large Enough to Be Useful?	36
	Confidence Intervals by the Chart Method	36
	Confidence Intervals by Fisher's Approximation	38
	Confidence Intervals by Ruben's Approximation	39
	Variable Groupings Based on Correlations	40
	Relationship to Factor Analysis	46
<b>2.2</b>	<b>Summary</b>	<b>46</b>
	Exercises	47

## 3. MULTIVARIATE DATA PLOTS

<b>3.1</b>	<b>Three-Dimensional Data Plots</b>	<b>55</b>
<b>3.2</b>	<b>Plots of Higher Dimensional Data</b>	<b>59</b>
	Chernoff Faces	61
	Star Plots and Sun-Ray Plots	63

	Andrews' Plots	65	
	Side-by-Side Scatter Plots	66	
<b>3.3</b>	<b>Plotting to Check for Multivariate Normality</b>	<b>67</b>	
	Summary	73	
	Exercises	73	
<b>4.</b>	<b>EIGENVALUES AND EIGENVECTORS</b>		<b>77</b>
<b>4.1</b>	<b>Trace and Determinant</b>	<b>77</b>	
	Examples	78	
<b>4.2</b>	<b>Eigenvalues</b>	<b>78</b>	
<b>4.3</b>	<b>Eigenvectors</b>	<b>79</b>	
	Positive Definite and Positive Semidefinite Matrices	80	
<b>4.4</b>	<b>Geometric Descriptions (<math>p = 2</math>)</b>	<b>82</b>	
	Vectors	82	
	Bivariate Normal Distributions	83	
<b>4.5</b>	<b>Geometric Descriptions (<math>p = 3</math>)</b>	<b>87</b>	
	Vectors	87	
	Trivariate Normal Distributions	87	
<b>4.6</b>	<b>Geometric Descriptions (<math>p &gt; 3</math>)</b>	<b>90</b>	
	Summary	91	
	Exercises	91	
<b>5.</b>	<b>PRINCIPAL COMPONENTS ANALYSIS</b>		<b>93</b>
<b>5.1</b>	<b>Reasons for Using Principal Components Analysis</b>	<b>93</b>	
	Data Screening	93	
	Clustering	95	
	Discriminant Analysis	95	
	Regression	95	
<b>5.2</b>	<b>Objectives of Principal Components Analysis</b>	<b>96</b>	
<b>5.3</b>	<b>Principal Components Analysis on the Variance-Covariance Matrix <math>\Sigma</math></b>	<b>96</b>	
	Principal Component Scores	98	
	Component Loading Vectors	98	



<b>5.4</b>	<b>Estimation of Principal Components</b>	<b>99</b>	
	Estimation of Principal Component Scores	99	
<b>5.5</b>	<b>Determining the Number of Principal Components</b>	<b>99</b>	
	Method 1	100	
	Method 2	100	
<b>5.6</b>	<b>Caveats</b>	<b>107</b>	
<b>5.7</b>	<b>PCA on the Correlation Matrix P</b>	<b>109</b>	
	Principal Component Scores	110	
	Component Correlation Vectors	110	
	Sample Correlation Matrix	110	
	Determining the Number of Principal Components	110	
<b>5.8</b>	<b>Testing for Independence of the Original Variables</b>	<b>111</b>	
<b>5.9</b>	<b>Structural Relationships</b>	<b>111</b>	
<b>5.10</b>	<b>Statistical Computing Packages</b>	<b>112</b>	
	SAS <sup>R</sup> PRINCOMP Procedure	112	
	Principal Components Analysis Using Factor Analysis Programs	118	
	PCA with SPSS's FACTOR Procedure	124	
	Summary	142	
	Exercises	142	
<b>6.</b>	<b>FACTOR ANALYSIS</b>		<b>14</b>
<b>6.1</b>	<b>Objectives of Factor Analysis</b>	<b>147</b>	
<b>6.2</b>	<b>Caveats</b>	<b>148</b>	
<b>6.3</b>	<b>Some History of Factor Analysis</b>	<b>148</b>	
<b>6.4</b>	<b>The Factor Analysis Model</b>	<b>150</b>	
	Assumptions	150	
	Matrix Form of the Factor Analysis Model	151	
	Definitions of Factor Analysis Terminology	151	
<b>6.5</b>	<b>Factor Analysis Equations</b>	<b>151</b>	
	Nonuniqueness of the Factors	152	
<b>6.6</b>	<b>Solving the Factor Analysis Equations</b>	<b>153</b>	

<b>6.7</b>	<b>Choosing the Appropriate Number of Factors</b>	<b>155</b>
	Subjective Criteria	156
	Objective Criteria	156
<b>6.8</b>	<b>Computer Solutions of the Factor Analysis Equations</b>	<b>157</b>
	Principal Factor Method on <b>R</b>	158
	Principal Factor Method with Iteration	159
<b>6.9</b>	<b>Rotating Factors</b>	<b>170</b>
	Examples ( $m = 2$ )	171
	Rotation Methods	172
	The Varimax Rotation Method	173
<b>6.10</b>	<b>Oblique Rotation Methods</b>	<b>174</b>
<b>6.11</b>	<b>Factor Scores</b>	<b>180</b>
	Bartlett's Method or the Weighted Least-Squares Method	181
	Thompson's Method or the Regression Method	181
	Ad Hoc Methods	181
	Summary	212
	Exercises	213
<b>7.</b>	<b>DISCRIMINANT ANALYSIS</b>	<b>217</b>
<b>7.1</b>	<b>Discrimination for Two Multivariate Normal Populations</b>	<b>217</b>
	A Likelihood Rule	218
	The Linear Discriminant Function Rule	218
	A Mahalanobis Distance Rule	218
	A Posterior Probability Rule	218
	Sample Discriminant Rules	219
	Estimating Probabilities of Misclassification	220
	Resubstitution Estimates	220
	Estimates from Holdout Data	220
	Cross-Validation Estimates	221
<b>7.2</b>	<b>Cost Functions and Prior Probabilities (Two Populations)</b>	<b>229</b>
<b>7.3</b>	<b>A General Discriminant Rule (Two Populations)</b>	<b>231</b>
	A Cost Function	232
	Prior Probabilities	232

	Average Cost of Misclassification	232
	A Bayes Rule	233
	Classification Functions	233
	Unequal Covariance Matrices	233
	Tricking Computing Packages	234
<b>7.4</b>	<b>Discriminant Rules (More than Two Populations)</b>	<b>235</b>
	Basic Discrimination	238
<b>7.5</b>	<b>Variable Selection Procedures</b>	<b>245</b>
	Forward Selection Procedure	245
	Backward Elimination Procedure	246
	Stepwise Selection Procedure	246
	Recommendations	247
	Caveats	247
<b>7.6</b>	<b>Canonical Discriminant Functions</b>	<b>255</b>
	The First Canonical Function	256
	A Second Canonical Function	257
	Determining the Dimensionality of the Canonical Space	260
	Discriminant Analysis with Categorical Predictor Variables	271
<b>7.7</b>	<b>Nearest Neighbor Discriminant Analysis</b>	<b>275</b>
<b>7.8</b>	<b>Classification Trees</b>	<b>283</b>
	Summary	283
	Exercises	283
<b>8.</b>	<b>LOGISTIC REGRESSION METHODS</b>	<b>288</b>
<b>8.1</b>	<b>Logistic Regression Model</b>	<b>287</b>
<b>8.2</b>	<b>The Logit Transformation</b>	<b>287</b>
	Model Fitting	288
<b>8.3</b>	<b>Variable Selection Methods</b>	<b>296</b>
<b>8.4</b>	<b>Logistic Discriminant Analysis (More Than Two Populations)</b>	<b>301</b>
	Logistic Regression Models	301
	Model Fitting	302
	Another SAS LOGISTIC Analysis	314
	Exercises	316

<b>9. CLUSTER ANALYSIS</b>	<b>319</b>
<b>9.1 Measures of Similarity and Dissimilarity</b>	<b>319</b>
Ruler Distance	319
Standardized Ruler Distance	320
A Mahalanobis Distance	320
Dissimilarity Measures	320
<b>9.2 Graphical Aids in Clustering</b>	<b>321</b>
Scatter Plots	321
Using Principal Components	322
Andrews' Plots	322
Other Methods	322
<b>9.3 Clustering Methods</b>	<b>322</b>
Nonhierarchical Clustering Methods	323
Hierarchical Clustering	323
Nearest Neighbor Method	323
A Hierarchical Tree Diagram	325
Other Hierarchical Clustering Methods	326
Comparisons of Clustering Methods	327
Verification of Clustering Methods	327
How Many Clusters?	327
Beale's $F$ -Type Statistic	328
A Pseudo Hotelling's $T^2$ Test	329
The Cubic Clustering Criterion	329
Clustering Order	334
Estimating the Number of Clusters	339
Principal Components Plots	348
Clustering with SPSS	355
SAS's FASTCLUS Procedure	369
<b>9.4 Multidimensional Scaling</b>	<b>385</b>
Exercises	395
<b>10. MEAN VECTORS AND VARIANCE-COVARIANCE MATRICES</b>	<b>397</b>
<b>10.1 Inference Procedures for Variance-Covariance Matrices</b>	<b>397</b>
A Test for a Specific Variance-Covariance Matrix	398
A Test for Sphericity	400

	A Test for Compound Symmetry	403
	A Test for the Huynh–Feldt Conditions	405
	A Test for Independence	406
	A Test for Independence of Subsets of Variables	407
	A Test for the Equality of Several Variance–Covariance Matrices	408
<b>10.2</b>	<b>Inference Procedures for a Mean Vector</b>	<b>408</b>
	Hotelling’s $T^2$ Statistic	409
	Hypothesis Test for $\mu$	409
	Confidence Region for $\mu$	409
	A More General Result	411
	Special Case—A Test of Symmetry	412
	A Test for Linear Trend	418
	Fitting a Line to Repeated Measures	418
	Multivariate Quality Control	419
<b>10.3</b>	<b>Two Sample Procedures</b>	<b>420</b>
	Repeated Measures Experiments	420
<b>10.4</b>	<b>Profile Analyses</b>	<b>431</b>
<b>10.5</b>	<b>Additional Two-Group Analyses</b>	<b>432</b>
	Paired Samples	432
	Unequal Variance–Covariance Matrices	433
	Large Sample Sizes	433
	Small Sample Sizes	433
	Summary	434
	Exercises	434
<b>11.</b>	<b>MULTIVARIATE ANALYSIS OF VARIANCE</b>	<b>4</b>
<b>11.1</b>	<b>MANOVA</b>	<b>439</b>
	MANOVA Assumptions	440
	Test Statistics	440
	Test Comparisons	441
	Why Do We Use MANOVAs?	441
	A Conservative Approach to Multiple Comparisons	442

11.2	<b>Dimensionality of the Alternative Hypothesis</b>	<b>455</b>
11.3	<b>Canonical Variates Analysis</b>	<b>456</b>
	The First Canonical Variate	456
	The Second Canonical Variate	457
	Other Canonical Variates	457
11.4	<b>Confidence Regions for Canonical Variates</b>	<b>458</b>
	Summary	485
	Exercises	485
<b>12</b>	<b>PREDICTION MODELS AND MULTIVARIATE REGRESSION</b>	<b>489</b>
12.1	<b>Multiple Regression</b>	<b>489</b>
12.2	<b>Canonical Correlation Analysis</b>	<b>494</b>
	Two Sets of Variables	494
	The First Canonical Correlation	495
	The Second Canonical Correlation	495
	Number of Canonical Correlations	496
	Estimates	496
	Hypothesis Tests on the Canonical Correlations	497
	Interpreting Canonical Functions	508
	Canonical Correlation Analysis with SPSS	511
12.3	<b>Factor Analysis and Regression</b>	<b>515</b>
	Summary	522
	Exercises	522
	<b>APPENDIX A: MATRIX RESULTS</b>	<b>525</b>
A.1	<b>Basic Definitions and Rules of Matrix Algebra</b>	<b>525</b>
A.2	<b>Quadratic Forms</b>	<b>527</b>
A.3	<b>Eigenvalues and Eigenvectors</b>	<b>528</b>
A.4	<b>Distances and Angles</b>	<b>529</b>
A.5	<b>Miscellaneous Results</b>	<b>529</b>

<b>APPENDIX B: WORK ATTITUDES SURVEY</b>	<b>5</b>
<b>B.1 Data File Structure</b>	<b>536</b>
<b>B.2 SPSS Data Entry Commands</b>	<b>538</b>
<b>B.3 SAS Data Entry Commands</b>	<b>543</b>
<b>APPENDIX C: FAMILY CONTROL STUDY</b>	<b>5</b>
<b>REFERENCES</b>	<b>5</b>
<b>Index</b>	<b>563</b>

# Applied Multivariate Methods

# 1

Multivariate data occur in all branches of science. Almost all data collected by today's researchers can be classified as multivariate data. For example, a marketing researcher might be interested in identifying characteristics of individuals that would enable the researcher to determine whether a certain individual is likely to purchase a specific product. Furthermore, a wheat breeder might be interested in more than just the yields of some new varieties of wheat. The wheat breeder may also be interested in these varieties' resistance to insect damage and drought. Finally, a social scientist might be interested in studying the relationships between teenage girls' dating behaviors and their fathers' attitudes. Each of these endeavors involves multivariate data.

To begin a discussion of multivariate data analysis methods, the concept of an experimental unit must be defined. An *experimental unit* is any object or item that can be measured or evaluated in some way. Measuring and evaluating experimental units is a principal activity of most researchers. Examples of experimental units include people, animals, insects, fields, plots of land, companies, trees, wheat kernels, and countries. *Multivariate data* result whenever a researcher measures or evaluates more than one attribute or characteristic of each experimental unit. These attributes or characteristics are usually called *variables* by statisticians.

The next section gives an overview of some multivariate methods that are discussed in this text.

## 1.1

### An Overview of Multivariate Methods

Multivariate methods are extremely useful for helping researchers make sense of large, complicated, and complex data sets that consist of a lot of variables measured on large numbers of experimental units. The importance and usefulness of multivariate methods increase as the number of variables being measured and the number of experimental units being evaluated increase.

Often, the primary objective of multivariate analyses is to summarize large amounts of data by means of relatively few parameters. The underlying theme behind many multivariate techniques is simplification.

Multivariate analyses are often concerned with finding relationships among (1) the response variables, (2) the experimental units, and (3) both response variables and experimental units. One might say that relationships exist among the response variables when several of the variables really are measuring a common entity. For example, suppose one gives tests to third-graders in reading, spelling, arithmetic, and science. Individual students may tend to get high scores, medium scores, or low scores in all four areas. If this did happen, then these tests would be related to one another. In such a case, the common thing that these tests may be measuring might be "overall intelligence."

Relationships might exist between the experimental units if some of them are similar to each other. For example, suppose breakfast cereals are evaluated for their nutritional content. One might measure the grams of fat, protein,



carbohydrates, and sodium in each cereal. Cereals would be related to each other if they tended to be similar with respect to the amounts of fat, protein carbohydrates, and sodium that are in a single serving of each cereal. One might expect sweetened cereals to be related to each other and high-fiber cereals to be related to each other. One might also expect sweetened cereal to be much different than high-fiber cereals.

Many multivariate techniques tend to be exploratory in nature rather than confirmatory. That is, many multivariate methods tend to motivate hypotheses rather than test them. Consider a situation in which a researcher may have 50 variables measured on more than 2000 experimental units. Traditional statistical methods usually require that a researcher state some hypotheses, collect some data, and then use these data to either substantiate or repudiate the hypotheses. An alternative situation that often exists is a case in which a researcher has a large amount of data available and wonders whether there might be valuable information in these data. Multivariate techniques are often useful for exploring data in an attempt to learn if there is worthwhile and valuable information contained in these data.

## Variable- and Individual-Directed Techniques

One fundamental distinction between multivariate methods is that some are classified as “variable-directed” techniques, while others are classified as “individual-directed” techniques.

*Variable-directed* techniques are those that are primarily concerned with relationships that might exist among the response variables being measured. Some examples of this type of technique are analyses performed on correlation matrices, principal components analysis, factor analysis, regression analysis, and canonical correlation analysis.

*Individual-directed* techniques are those that are primarily concerned with relationships that might exist among the experimental units and/or individuals being measured. Some examples of this type of technique are discriminant analysis, cluster analysis, and multivariate analysis of variance (MANOVA).

## Creating New Variables

We quite often find it useful to create new variables for each experimental unit so they can be compared to each other more easily. Many multivariate methods help researchers create new variables that have desirable properties.

Some of the multivariate techniques that create new variables are principal components analysis, factor analysis, canonical correlation analysis, canonical discriminant analysis, and canonical variates analysis.

Some brief overviews of the multivariate techniques that are considered in this book are given next.