

外国优秀统计学教材系列丛书 · 影印版

PEARSON  
Prentice  
Hall

作者:

Richard A. Johnson

Dean W. Wichern

中国统计出版社  
China Statistics Press



全国统计教材编审委员会组织引进

# Applied Multivariate Statistical Analysis (5th Ed)

应用多元统计分析

Pearson  
Education





# 应用多元统计分析

## Applied Multivariate Statistical Analysis(5th Ed)

RICHARD A. JOHNSON DEAN W. WICHERN



英文版  
附光盘



B1271652



中国统计出版社  
China Statistics Press

Applied Multivariate Statistical Analysis

RICHARD A. JOHNSON DEAN W. WICHERN

ISBN 0-13-092553-5

Copyright © 2003. Exclusive rights by Pearson Education, Inc. for manufacture and export. This book cannot be re-exported from the country to which it is sold by Pearson Education.

Original language published by Pearson Education, Inc. All Rights reserved. No part of this publication may be reproduced or distributed in any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

Authorized English language reprint edition jointly published by Pearson Education and China Statistics Press. For sale and distribution in the People's Republic of China exclusively (except Taiwan, Hong Kong SAR and Macau SAR).

本书英文影印版由中国统计出版社和美国培生教育出版集团合作出版。此版本仅限于中华人民共和国境内（不包括中国香港、澳门特别行政区和中国台湾地区）销售发行。

本书封面贴有 Pearson Education（培生教育出版集团）激光防伪标签。无标签者不得销售。

北京市版权局著作权合同登记号：01-2003-3141

## (京)新登字 041 号

图书在版编目(CIP)数据

### 应用多元统计分析

Applied Multivariate Statistical Analysis

/ (美) 约翰逊(RICHARD A. JOHNSON) (美) 威克瑞恩(DEAN W. WICHERN) 著 - 影印本

北京: 中国统计出版社, 2003. 5

ISBN 7-5037-4099-X

I. 应… II. ①约… ②威 III. 多元分析: 统计分析 - 英文 IV. 0212.4

中国版本图书馆CIP数据核字(2003)第039016号

责任编辑 / 刘国宁 吕 军

封面设计 / 刘国宁 张建民

E-mail / cbsebs@stats.gov.cn

出版发行 / 中国统计出版社

通信地址 / 北京市西城区月坛南街75号

邮政编码 / 100826

办公地址 / 北京市丰台区西三环南路甲6号

电 话 / (010) 63459084 63266600-22500(发行部)

印 刷 / 北京市顺义兴华印刷厂

经 销 / 新华书店

开 本 / 850 × 1092 毫米 1/16

字 数 / 1420 千字

印 张 / 49.5

印 数 / 1-3000 册

版 别 / 2003年5月第1版

版 次 / 2003年5月第1次印刷

书 号 / ISBN 7-5037-4099-X/0.46

定 价 / 68.00 元

中国统计版图书, 如有印装错误, 本社发行部负责调换。

## 引进外国优秀统计学教材专家委员会

**主任:** 贺 铿 国家统计局副局长

**副主任:** 王吉利 国家统计局统计教育中心主任  
谢鸿光 中国统计出版社社长

**委员:** (按姓氏笔画排序)

文兼武	国家统计局统计科学研究所	所长
孙山泽	北京大学数学科学学院概率统计系	教授
严建辉	中国统计出版社	副总编
吴喜之	中国人民大学统计学系	教授
何书元	北京大学数学科学学院概率统计系	教授
张尧庭	上海财经大学经济学院	教授
张润楚	南开大学数学学院统计学系	教授
邱 东	东北财经大学	教授
陈 江	美国雪城大学管理学院	教授
陈家鼎	北京大学数学科学学院概率统计系	教授
杨振海	北京工业大学应用数理学院	教授
孟晓犁	哈佛大学统计系	教授
林正炎	浙江大学理学院数学系	教授
茆诗松	华东师范大学统计系	教授
郑 明	复旦大学管理学院统计系	副教授
柯惠新	北京广播学院	教授
徐国祥	上海财经大学统计学系	教授
袁 卫	中国人民大学	教授
蒋 萍	东北财经大学统计系	教授
谢邦昌	台湾辅仁大学统计资讯学系	教授
濮晓龙	华东师范大学统计系	副教授

**办公室:**

温 明	国家统计局统计教育中心教材处	处长
刘国宁	中国统计出版社第二书籍编辑部	主任
孙洪娟	国家统计局统计教育中心教材处	统计师
吕 军	中国统计出版社第二书籍编辑部	编辑

## 出版说明

21

世纪的竞争是人才的竞争，是全球性人才培养机制的较量。

如何培养面向现代化、面向世界、面向未来的高素质的人才成为我国人才培养的当务之急。为此，教育部发出通知，倡导在全国普通高等学校中使用原版外国教材，进行双语教学，培养适应经济全球化的人才。

为了响应教育部的号召，促进统计教材的改革，培养既懂统计专业知识又具备较高英语语言能力的统计人才，全国统计教材编审委员会在国家统计局领导的大力支持下，组织引进了这套“外国优秀统计学教材”。

为了做好“外国优秀统计学教材”引进工作，全国统计教材编审委员会将其列入了“十五”规划，并成立了由海内外统计学家组成的专家委员会。在对国外统计学教材的使用情况进行了充分了解，对国内高等院校使用外国统计学教材的需求情况进行了仔细分析，并对从各种渠道推荐来的统计教材进行了认真审定的基础上，制定了引进教材书目。在确定引进教材书目的过程中，我们得到了国内外有关专家、有关院校和外国出版公司及其北京办事处的支持和帮助，在此致谢。中国人民大学统计学系的吴喜之教授不仅推荐了大量的优秀候选书目，而且校译了影印教材的翻译目录，为这套教材的及早出版作了大量的工作，我们表示衷心的感谢。

这套引进教材多数是国外再版多次、反响良好，又比较适合国内情况、易于教学的统计教材。我们希望这套引进教材的出版对促进我国统计教材的改革和高校统计学专业双语教学的发展能够起到重要的推动作用。

全国统计教材编审委员会

2002年8月28日

*To the memory of my mother and my father.*

*R. A. J.*

*To Dorothy, Michael, and Andrew.*

*D. W. W.*

---

# Preface

## INTENDED AUDIENCE

This book originally grew out of our lecture notes for an “Applied Multivariate Analysis” course offered jointly by the Statistics Department and the School of Business at the University of Wisconsin–Madison. *Applied Multivariate Statistical Analysis*, Fifth Edition, is concerned with statistical methods for describing and analyzing multivariate data. Data analysis, while interesting with one variable, becomes truly fascinating and challenging when several variables are involved. Researchers in the biological, physical, and social sciences frequently collect measurements on several variables. Modern computer packages readily provide the numerical results to rather complex statistical analyses. We have tried to provide readers with the supporting knowledge necessary for making proper interpretations, selecting appropriate techniques, and understanding their strengths and weaknesses. We hope our discussions will meet the needs of experimental scientists, in a wide variety of subject matter areas, as a readable introduction to the statistical analysis of multivariate observations.

## LEVEL

Our aim is to present the concepts and methods of multivariate analysis at a level that is readily understandable by readers who have taken two or more statistics courses. We emphasize the applications of multivariate methods and, consequently, have attempted to make the mathematics as palatable as possible. We avoid the use of calculus. On the other hand, the concepts of a matrix and of matrix manipulations are important. We do not assume the reader is familiar with matrix algebra. Rather, we introduce matrices as they appear naturally in our discussions, and we then show how they simplify the presentation of multivariate models and techniques.

The introductory account of matrix algebra, in Chapter 2, highlights the more important matrix algebra results as they apply to multivariate analysis. The Chapter 2 supplement provides a summary of matrix algebra results for those with little or no previous exposure to the subject. This supplementary material helps make the book self-contained and is used to complete proofs. The proofs may be ignored on the first reading. In this way we hope to make the book accessible to a wide audience.

In our attempt to make the study of multivariate analysis appealing to a large audience of both practitioners and theoreticians, we have had to sacrifice a consistency

of level. Some sections are harder than others. In particular, we have summarized a voluminous amount of material on regression in Chapter 7. The resulting presentation is rather succinct and difficult the first time through. We hope instructors will be able to compensate for the unevenness in level by judiciously choosing those sections, and subsections, appropriate for their students and by toning them down if necessary.

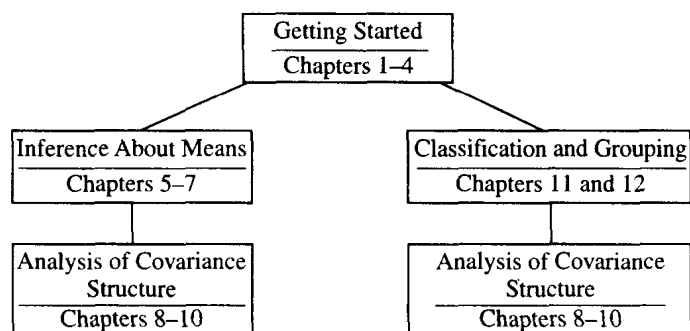
## ORGANIZATION AND APPROACH

The methodological “tools” of multivariate analysis are contained in Chapters 5 through 12. These chapters represent the heart of the book, but they cannot be assimilated without much of the material in the introductory Chapters 1 through 4. Even those readers with a good knowledge of matrix algebra or those willing to accept the mathematical results on faith should, at the very least, peruse Chapter 3, “Sample Geometry,” and Chapter 4, “Multivariate Normal Distribution.”

Our approach in the methodological chapters is to keep the discussion direct and uncluttered. Typically, we start with a formulation of the population models, delineate the corresponding sample results, and liberally illustrate everything with examples. The examples are of two types: those that are simple and whose calculations can be easily done by hand, and those that rely on real-world data and computer software. These will provide an opportunity to (1) duplicate our analyses, (2) carry out the analyses dictated by exercises, or (3) analyze the data using methods other than the ones we have used or suggested.

The division of the methodological chapters (5 through 12) into three units allows instructors some flexibility in tailoring a course to their needs. Possible sequences for a one-semester (two quarter) course are indicated schematically.

Each instructor will undoubtedly omit certain sections from some chapters to cover a broader collection of topics than is indicated by these two choices.



For most students, we would suggest a quick pass through the first four chapters (concentrating primarily on the material in Chapter 1; Sections 2.1, 2.2, 2.3, 2.5, 2.6, and 3.6; and the “assessing normality” material in Chapter 4) followed by a selection of methodological topics. For example, one might discuss the comparison of mean vectors, principal components, factor analysis, discriminant analysis and clustering. The discussions could feature the many “worked out” examples included in



these sections of the text. Instructors may rely on diagrams and verbal descriptions to teach the corresponding theoretical developments. If the students have uniformly strong mathematical backgrounds, much of the book can successfully be covered in one term.

We have found individual data-analysis projects useful for integrating material from several of the methods chapters. Here, our rather complete treatments of multivariate analysis of variance (MANOVA), regression analysis, factor analysis, canonical correlation, discriminant analysis, and so forth are helpful, even though they may not be specifically covered in lectures.

## CHANGES TO THE FIFTH EDITION

**New material.** Users of the previous editions will notice that we have added several exercises and data sets, some new graphics, and have expanded the discussion of the dimensionality of multivariate data, growth curves and classification and regression trees (CART). In addition, the algebraic development of correspondence analysis has been redone and a new section on data mining has been added to Chapter 12. We put the data mining material in Chapter 12 since much of data mining, as it is now applied in business, has a classification and/or grouping objective. As always, we have tried to improve the exposition in several places.

**Data CD.** Recognizing the importance of modern statistical packages in the analysis of multivariate data, we have added numerous real-data sets. The full data sets used in the book are saved as ASCII files on the CD-ROM that is packaged with each copy of the book. This format will allow easy interface with existing statistical software packages and provide more convenient hands-on data analysis opportunities.

**Instructors Solutions Manual.** An *Instructors Solutions Manual* (ISBN 0-13-092555-1) containing complete solutions to most of the exercises in the book is available free upon adoption from Prentice Hall.

For information on additional for sale supplements that may be used with the book or additional titles of interest, please visit the Prentice Hall Web site at [www.prenhall.com](http://www.prenhall.com).

## ACKNOWLEDGMENTS

We thank our many colleagues who helped improve the applied aspect of the book by contributing their own data sets for examples and exercises. A number of individuals helped guide this revision, and we are grateful for their suggestions: Steve Coad, University of Michigan; Richard Kiltie, University of Florida; Sam Kotz, George Mason University; Shyamal Peddada, University of Virginia; K. Sivakumar, University of Illinois at Chicago; Eric Smith, Virginia Tech; and Stanley Wasserman, University of Illinois at Urbana-Champaign. We also acknowledge the feedback of the students we have taught these past 30 years in our applied multivariate analysis courses. Their comments and suggestions are largely responsible for the present iteration

of this work. We would also like to give special thanks to Wai Kwong Cheang for his help with the calculations for many of the examples.

We must thank Dianne Hall for her valuable work on the CD-ROM and Solutions Manual, Steve Verrill for computing assistance throughout, and Alison Pollack for implementing a Chernoff faces program. We are indebted to Cliff Gilman for his assistance with the multidimensional scaling examples discussed in Chapter 12. Jacquelyn Forer did most of the typing of the original draft manuscript, and we appreciate her expertise and willingness to endure the cajoling of authors faced with publication deadlines. Finally, we would like to thank Quincy McDonald, Joanne Wendelken, Steven Scott Pawlowski, Pat Daly, Linda Behrens, Alan Fischer, and the rest of the Prentice Hall staff for their help with this project.

*R. A. Johnson*  
rich@stat.wisc.edu

*D. W. Wichern*  
d-wichern@tamu.edu

# 目录

## 前言

- 1 多元分析的各个方面 1
  - 1.1 介绍 1
  - 1.2 多元方法的应用 3
  - 1.3 组织数据 5
  - 1.4 数据的显示和图表示 19
  - 1.5 距离 30
  - 1.6 最终说明 38
- 2 矩阵代数和随机向量 50
  - 2.1 介绍 50
  - 2.2 矩阵和向量代数的一些基础 50
  - 2.3 正定矩阵 61
  - 2.4 平方根矩阵 66
  - 2.5 随机向量和矩阵 67
  - 2.6 均值向量和协方差矩阵 68
  - 2.7 矩阵不等式和最大化 79
- 3 样本几何学和随机抽样 112
  - 3.1 介绍 112
  - 3.2 样本几何学 112
  - 3.3 随机样本和样本平均与协方差矩阵的期望值 120
  - 3.4 广义方差 124
  - 3.5 作为矩阵运算形式的样本平均、方差和相关 139
  - 3.6 变量线性组合的样本值 141

4	多元正态分布	149
4.1	介绍	149
4.2	多元变量正态密度及其性质	149
4.3	多元正态分布的抽样和最大似然估计	168
4.4	$\bar{X}$ 和S的抽样分布	173
4.5	$\bar{X}$ 和S的大样本性质	175
4.6	评估正态性假定	177
4.7	检测离群值和清理数据	189
4.8	变换到近似正态	194
5	关于均值向量的推断	210
5.1	介绍	210
5.2	$\mu_0$ 作为正态总体均值的合理性	210
5.3	HotellingT <sup>2</sup> 和似然比检验	216
5.4	置信区域和成分均值的同时比较	220
5.5	有关一个总体均值向量的大样本推断	234
5.6	多元质量控制图	239
5.7	当有缺失值时的关关均值向量的推断	252
5.8	由于多变量观测值依赖于时间而造成的困难	256
6	几个多元均值的比较	272
6.1	介绍	272
6.2	配对比较和重复度量设计	272
6.3	二总体的均值向量的比较	283
6.4	比较几个多元总体均值(单向Manova)	293
6.5	处理效应的同时置信区间	305
6.6	双向多元方差分析	307
6.7	纵断面分析	318
6.8	重复测量设计和增长曲线	323
6.9	分析多元变量模型的看法和策略	327
7	多元线性回归模型	354
7.1	介绍	354
7.2	经典线性回归模型	354
7.3	最小二乘估计	358
7.4	关于回归模型的推断	365
7.5	基于估计回归函数的推断	374
7.6	模型的核对和回归的其他方面	377

- 7.7 多元多重回归 383
- 7.8 线性回归的概念 398
- 7.9 比较回归模型的两种公式形式 407
- 7.10 带有时间相依误差的多重回归模型 410
- 8 主成份 426**
  - 8.1 介绍 426
  - 8.2 总体主成份 426
  - 8.3 由主成份概括样本变异 437
  - 8.4 绘制主成份图 450
  - 8.5 大样本推断 452
  - 8.6 用主成份监督质量 455
- 9 因子分析和对结构协方差矩阵的推断 477**
  - 9.1 介绍 477
  - 9.2 正交因子模型 478
  - 9.3 估计方法 484
  - 9.4 因子旋转 501
  - 9.5 因子得分 510
  - 9.6 因子分析的看法和策略 517
  - 9.7 结构方程模型 524
- 10 典型相关分析 543**
  - 10.1 介绍 543
  - 10.2 典型变量和典型相关 543
  - 10.3 解释总体典型变量 551
  - 10.4 样本典型变量和样本典型相关 556
  - 10.5 附加样本描述性度量 564
  - 10.6 大样本推断 569
- 11 判别和分类 581**
  - 11.1 介绍 581
  - 11.2 两总体的分离和分类 582
  - 11.3 二多元正态总体的分类 590
  - 11.4 估计分类函数 598
  - 11.5 费歇尔判别函数——总体分离 609
  - 11.6 多总体分类 612
  - 11.7 关于几个总体的费歇尔的判别方法 628



11.8 最后评论 641

**12 聚类、距离法和分级 668**

12.1 介绍 668

12.2 相似测度 670

12.3 分层聚类法 679

12.4 非分层聚类法 694

12.5 多维尺度 700

12.6 对应分析 709

12.7 观察抽样单位和变量的双图 719

12.8 Procrustes 分析:比较结构的一种方法 723

**附录 748**

**数据索引 758**

**主题索引 761**

---

---

# Contents

<b>PREFACE</b>	<b>xv</b>
<b>1 ASPECTS OF MULTIVARIATE ANALYSIS</b>	<b>1</b>
1.1 Introduction	1
1.2 Applications of Multivariate Techniques	3
1.3 The Organization of Data	5
<i>Arrays, 5</i>	
<i>Descriptive Statistics, 6</i>	
<i>Graphical Techniques, 11</i>	
1.4 Data Displays and Pictorial Representations	19
<i>Linking Multiple Two-Dimensional Scatter Plots, 20</i>	
<i>Graphs of Growth Curves, 24</i>	
<i>Stars, 25</i>	
<i>Chernoff Faces, 28</i>	
1.5 Distance	30
1.6 Final Comments	38
Exercises	38
References	48
<b>2 MATRIX ALGEBRA AND RANDOM VECTORS</b>	<b>50</b>
2.1 Introduction	50
2.2 Some Basics of Matrix and Vector Algebra	50
<i>Vectors, 50</i>	
<i>Matrices, 55</i>	
2.3 Positive Definite Matrices	61
2.4 A Square-Root Matrix	66
2.5 Random Vectors and Matrices	67
2.6 Mean Vectors and Covariance Matrices	68
<i>Partitioning the Covariance Matrix, 74</i>	
<i>The Mean Vector and Covariance Matrix</i>	
<i>for Linear Combinations of Random Variables, 76</i>	
<i>Partitioning the Sample Mean Vector</i>	
<i>and Covariance Matrix, 78</i>	
2.7 Matrix Inequalities and Maximization	79

Supplement 2A: Vectors and Matrices: Basic Concepts	84
<i>Vectors</i> ,	84
<i>Matrices</i> ,	89
Exercises	104
References	111

### **3 SAMPLE GEOMETRY AND RANDOM SAMPLING** **112**

3.1	Introduction	112
3.2	The Geometry of the Sample	112
3.3	Random Samples and the Expected Values of the Sample Mean and Covariance Matrix	120
3.4	Generalized Variance	124
	<i>Situations in which the Generalized Sample Variance Is Zero</i> ,	130
	<i>Generalized Variance Determined by <math> R </math></i>	
	<i>and Its Geometrical Interpretation</i> ,	136
	<i>Another Generalization of Variance</i> ,	138
3.5	Sample Mean, Covariance, and Correlation As Matrix Operations	139
3.6	Sample Values of Linear Combinations of Variables	141
	Exercises	145
	References	148

### **4 THE MULTIVARIATE NORMAL DISTRIBUTION** **149**

4.1	Introduction	149
4.2	The Multivariate Normal Density and Its Properties	149
	<i>Additional Properties of the Multivariate Normal Distribution</i> ,	156
4.3	Sampling from a Multivariate Normal Distribution and Maximum Likelihood Estimation	168
	<i>The Multivariate Normal Likelihood</i> ,	168
	<i>Maximum Likelihood Estimation of <math>\mu</math> and <math>\Sigma</math></i> ,	170
	<i>Sufficient Statistics</i> ,	173
4.4	The Sampling Distribution of $\bar{\mathbf{X}}$ and $S$	173
	<i>Properties of the Wishart Distribution</i> ,	174
4.5	Large-Sample Behavior of $\bar{\mathbf{X}}$ and $S$	175
4.6	Assessing the Assumption of Normality	177
	<i>Evaluating the Normality of the Univariate Marginal Distributions</i> ,	178
	<i>Evaluating Bivariate Normality</i> ,	183
4.7	Detecting Outliers and Cleaning Data	189
	<i>Steps for Detecting Outliers</i> ,	190
4.8	Transformations To Near Normality	194
	<i>Transforming Multivariate Observations</i> ,	198
	Exercises	202
	References	209

<b>5</b>	<b>INFERENCES ABOUT A MEAN VECTOR</b>	<b>210</b>
5.1	Introduction	210
5.2	The Plausibility of $\mu_0$ as a Value for a Normal Population Mean	210
5.3	Hotelling's $T^2$ and Likelihood Ratio Tests	216
	<i>General Likelihood Ratio Method</i> ,	219
5.4	Confidence Regions and Simultaneous Comparisons of Component Means	220
	<i>Simultaneous Confidence Statements</i> ,	223
	<i>A Comparison of Simultaneous Confidence Intervals with One-at-a-Time Intervals</i> ,	229
	<i>The Bonferroni Method of Multiple Comparisons</i> ,	232
5.5	Large Sample Inferences about a Population Mean Vector	234
5.6	Multivariate Quality Control Charts	239
	<i>Charts for Monitoring a Sample of Individual Multivariate Observations for Stability</i> ,	241
	<i>Control Regions for Future Individual Observations</i> ,	247
	<i>Control Ellipse for Future Observations</i> ,	248
	<i><math>T^2</math>-Chart for Future Observations</i> ,	248
	<i>Control Charts Based on Subsample Means</i> ,	249
	<i>Control Regions for Future Subsample Observations</i> ,	251
5.7	Inferences about Mean Vectors when Some Observations Are Missing	252
5.8	Difficulties Due to Time Dependence in Multivariate Observations	256
	Supplement 5A: Simultaneous Confidence Intervals and Ellipses as Shadows of the p-Dimensional Ellipsoids	258
	Exercises	260
	References	270
<b>6</b>	<b>COMPARISONS OF SEVERAL MULTIVARIATE MEANS</b>	<b>272</b>
6.1	Introduction	272
6.2	Paired Comparisons and a Repeated Measures Design	272
	<i>Paired Comparisons</i> ,	272
	<i>A Repeated Measures Design for Comparing Treatments</i> ,	278
6.3	Comparing Mean Vectors from Two Populations	283
	<i>Assumptions Concerning the Structure of the Data</i> ,	283
	<i>Further Assumptions when <math>n_1</math> and <math>n_2</math> Are Small</i> ,	284
	<i>Simultaneous Confidence Intervals</i> ,	287
	<i>The Two-Sample Situation when <math>\Sigma_1 \neq \Sigma_2</math></i> ,	290
6.4	Comparing Several Multivariate Population Means (One-Way Manova)	293
	<i>Assumptions about the Structure of the Data for One-way MANOVA</i> ,	293
	<i>A Summary of Univariate ANOVA</i> ,	293
	<i>Multivariate Analysis of Variance (MANOVA)</i> ,	298