

Cambridge University Press
978-0-521-09070-4 - Topics in Applied Multivariate Analysis
Edited by Douglas M. Hawkins
Frontmatter
[More information](#)

Topics in Applied Multivariate Analysis

Edited by

DOUGLAS M. HAWKINS

Senior Consultant in Statistics

Council for Scientific and Industrial Research

CAMBRIDGE UNIVERSITY PRESS

CAMBRIDGE

LONDON NEW YORK NEW ROCHELLE

MELBOURNE SYDNEY

Cambridge University Press
978-0-521-09070-4 - Topics in Applied Multivariate Analysis
Edited by Douglas M. Hawkins
Frontmatter
[More information](#)

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo, Delhi

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521243681

© Cambridge University Press 1982

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 1982
This digitally printed version 2008

A catalogue record for this publication is available from the British Library

Library of Congress Catalogue Card Number: 81-15527

ISBN 978-0-521-24368-1 hardback
ISBN 978-0-521-09070-4 paperback

Cambridge University Press
978-0-521-09070-4 - Topics in Applied Multivariate Analysis
Edited by Douglas M. Hawkins
Frontmatter
[More information](#)

PREFACE

In 1980, the National Research Institute for Mathematical Sciences (NRIMS) of the Council for Scientific and Industrial Research held a highly successful Summer Seminar Series on Extensions of Linear-Quadratic Control Theory. This was followed in February 1981 by one on Applied Multivariate Analysis. The papers presented at that seminar formed the penultimate drafts of those in this book.

The choice of Applied Multivariate Analysis as the topic for the second Summer Seminar Series was prompted by the recognition that there is a much wider divergence between theory and practice of multivariate methods than is the case with univariate statistics. As a result of this, multivariate methods are applied widely, but little understood by their users.

The authors of these papers, apart from having considerable practical experience in the application of the techniques they discuss, have also contributed to their theoretical development. Their papers give a distillation of what they regard as the major features of each topic from the viewpoint of applicability, and so may go some way towards bridging the chasm between theory and practice.

Valuable comments on the first drafts and suggestions for improvements were made by Mr. J.C. Gower (Chapters 4 and 6), Professor J.N. Darroch (Chapter 3) and Professor J. Aitchison (Chapter 1). The generosity of these reviewers in finding time to go through the manuscript is greatly appreciated.

On behalf of the authors, I should like to acknowledge gratefully the support of NRIMS in subventing and hosting the Summer Seminar Series. The help of colleagues in our various institutions was also invaluable:- in particular, thanks are due to Mrs. M. Russouw (assisted by Mrs. A.M. de Villiers) for typing the manuscript; Mesdames P.Colman and A. Myburgh for preparation of graphics, and Mrs. J.S. Galpin for editorial assistance.

Douglas M. Hawkins
NRIMS, CSIR
1981.07.07

Cambridge University Press
978-0-521-09070-4 - Topics in Applied Multivariate Analysis
Edited by Douglas M. Hawkins
Frontmatter
[More information](#)

CONTRIBUTORS

Prof M W Browne
Dept of Statistics
University of South Africa
P O Box 392
PRETORIA 0001

Dr L P Fatti
National Research Institute
for Mathematical Sciences
C S I R
P O Box 395
PRETORIA 0001

Dr M J Greenacre
Dept of Statistics
University of South Africa
P O Box 392
PRETORIA 0001

Dr D M Hawkins
National Research Institute
for Mathematical Sciences
C S I R
P O Box 395
PRETORIA 0001

Prof G V Kass
Dept of Applied Mathematics
University of the Witwatersrand
Milner Park
JOHANNESBURG 2001

Dr T J v W Kotze
Institute for Biostatistics of
the SA Medical Research Council
P O Box 70
TYGERBERG 7505

Mr M W Muller
National Institute for Personnel
Research
P O Box 32410
BRAAMFONTEIN 2017

Dr E L Raath
National Research Institute
for Mathematical Sciences
C S I R
P O Box 395
PRETORIA 0001

Mrs J A ten Krooden
National Research Institute for
Mathematical Sciences
C S I R
P O Box 395
PRETORIA 0001

Dr L G Underhill
Dept of Statistics
University of Cape Town
RONDEBOSCH 7700

CONTENTS

CHAPTER 1	DISCRIMINANT ANALYSIS	1
	<i>L Paul Fatti, Douglas M Hawkins, E Liefde Raath</i>	
1.	PRELIMINARIES	1
1.1	General introduction	1
1.2	The basic principles of discriminant analysis	3
2.	CLASSICAL DISCRIMINANT ANALYSIS	5
2.1	Known parameters	5
2.1.1	The case of equal covariance matrices	6
2.1.2	Probability of misclassification	7
2.1.3	Canonical variables	8
2.2	Unknown parameters	9
2.2.1	Probabilities of misclassification (Equal Covariance Matrices)	11
2.2.2	The random effects model	16
2.2.3	Biased discriminant analysis	16
2.3	Robustness of the discriminant function	17
2.3.1	Use of dichotomous variables in the discriminant functions	18
2.3.2	Robustness to heteroscedasticity	19
2.3.3	More robust procedures	20
2.3.4	Discrimination using ranks	21
3.	VARIABLE SELECTION	24
3.1	Selection criteria	24
3.1.1	Heteroscedastic selection criterion	26
3.2	Variable selection procedures	27
4.	PREDICTIVE DISCRIMINATION	31
4.1	Introduction	31
4.2	Some results in predictive discrimination	31
4.2.1	General theory	31
4.2.2	Multivariate normal discrimination	32

4.3	Proportional covariance matrices	33
4.4	The random effects model	34
4.5	Predictive discrimination with an informative prior for the covariance matrix	35
4.6	Comparison of the predictive and classical approaches to discriminant analysis	36
5.	OTHER APPROACHES TO DISCRIMINANT ANALYSIS	38
5.1	Logistic discriminant analysis	39
5.2	Binary and continuous variables - the location model	41
5.3	Multinomial discriminant analysis	43
5.4	Discriminant analysis by the kernel method	45
5.5	Two methods based on ranks	48
6.	MISCELLANEOUS TOPICS	49
6.1	Sequential discrimination	49
6.2	Discriminant analysis using unclassified observations	51
6.3	Missing values in discriminant analysis	53
6.4	Testing for unsampled source populations	53
6.5	Constrained discriminant analysis	54
6.6	Equal mean discriminant analysis	55
7.	TESTING FOR MODEL FIT	55
	REFERENCES	65
	CHAPTER 2 COVARIANCE STRUCTURES	72
	<i>Michael W Browne</i>	
1.	INTRODUCTION	72
	PART I GENERAL TECHNICAL BACKGROUND	73
1.1	Genesis of structural models for covariance matrices	73
1.2	Invariance of covariance structures under changes of scale	75
1.3	Estimability of parameters	77
1.4	Covariance structures as regression models	79
1.5	Estimation and fit	80
1.6	Asymptotic distributions of estimates	90

1.7	Tests of fit	97
1.8	Selection of a model	101
1.9	Computational considerations	102
1.10	General Remarks	110
PART II SPECIFIC APPLICATIONS		112
2.1	Introductory comments	112
2.2	Linear covariance structures	112
2.3	Patterned correlation matrices	114
2.4	Unrestricted factor analysis	114
2.5	Multiple battery factor analysis	118
2.6	Nested models	118
2.7	Time series structures	120
2.8	Direct product structures	122
2.9	Some numerical examples	124
REFERENCES		133
CHAPTER 3 THE LOG-LINEAR MODEL AND ITS APPLICATION TO MULTI-WAY CONTINGENCY TABLES		142
<i>Theunis J v W Kotze</i>		
1.	INTRODUCTION	142
2.	NOTATION	143
3.	INTERACTION IN CONTINGENCY TABLES	144
4.	TWO TYPES OF INVESTIGATIONS	146
5.	THE MINIMUM DISCRIMINATION INFORMATION (MDI) APPROACH	146
5.1	Formulation of the MDI statistic	147
5.2	Partitioning of the test-statistic	150
6.	A GENERAL APPROACH VIA EXPONENTIAL DISTRIBUTIONS	151
6.1	Some properties of the exponential family of distributions	152
6.2	Fitting the generalized linear model	152
7.	STRATEGIES FOR SELECTING A FINAL LOG-LINEAR MODEL	154
8.	PROBLEMS OF APPLYING LOG-LINEAR MODELS TO MULTI-WAY CONTINGENCY TABLES	159
8.1	Aspects of sparse tables	159

Cambridge University Press
 978-0-521-09070-4 - Topics in Applied Multivariate Analysis
 Edited by Douglas M. Hawkins
 Frontmatter
[More information](#)

8.2	Outliers in Contingency Tables	160
8.3	Collapsing of tables	161
8.4	Distributional aspects	163
9.	COMPUTER PROGRAMS AVAILABLE	167
10.	THE APPLICATION OF LOG-LINEAR MODELS TO THE ANALYSIS OF RATES AND PROPORTIONS	167
10.1	Introduction	167
10.2	Logistic regression	169
11.	THE GRAPHICAL REPRESENTATION OF THE RELATIONSHIPS PRESENT IN A MULTI-WAY TABLE	171
11.1	Determination of associations present in the appropriate model	171
11.2	Graphical representation	173
11.3	Practical example	173
11.4	Discussion of graphical methods	180
12.	CONCLUSION	180
	REFERENCES	181
CHAPTER 4 SCALING A DATA MATRIX IN A LOW-DIMENSIONAL EUCLIDEAN SPACE		183
<i>Michael J Greenacre, Leslie G Underhill</i>		
1.	INTRODUCTION	183
2.	BASIC STRUCTURE DISPLAYS	186
2.1	Generalized basic structure	186
2.2	Basic structure displays	189
2.3	Examples and applications of basic structure displays	194
2.3.1	Principal components analysis and biplot	194
2.3.2	Symmetric biplot	198
2.3.3	Principal components with weights	200
2.3.4	Covariance biplot	201
2.3.5	Canonical variate analysis	205
2.3.6	Correspondence analysis	205
2.4	Stability of results and missing data	207

3.	CLASSICAL SCALING AND PRINCIPAL CO-ORDINATES ANALYSIS	208
3.1	Eigenstructure of a matrix of scalar products	208
3.2	Classical (Young-Householder) scaling	210
3.3	Principal co-ordinates analysis	210
3.4	When is a "distance matrix" a distance matrix?	211
3.5	Robustness of classical scaling	212
3.6	Examples of classical scaling	214
4.	LEAST SQUARES SCALING	219
4.1	Introduction	219
4.2	Examples of least squares scaling	226
5.	NONMETRIC SCALING	228
5.1	Monotonic regression	228
5.2	Kruskal-Shepard scaling	234
5.3	Alternative stress functions for nonmetric scaling	239
5.4	Local order nonmetric scaling	243
5.5	Local minima, starting configurations and choice of dimensionality	245
5.6	Incomplete data and replicated data	246
5.7	Examples of nonmetric scaling	247
6.	MISCELLANEOUS SCALING TOPICS	253
6.1	Interpretation of scaling output	253
6.2	Simulation studies	255
6.3	Comparison of solutions	256
6.4	Multidimensional unfolding	259
	REFERENCES	261
	CHAPTER 5 AUTOMATIC INTERACTION DETECTION	267
	<i>Douglas M Hawkins, Gordon V Kass</i>	
1.	INTRODUCTION	267
2.	MODE OF OPERATION	269
2.1	Types of predictor	271
2.2	Analysis of a set of cases	275
2.3	AID test statistics	276

3.	FORMAL INFERENCE IN AID	279
3.1	Exact and asymptotic theory	279
3.2	Conservative theory	280
3.3	Many predictors	283
3.4	The reduction to k categories	283
4.	OUTPUT FROM AN AID	285
5.	USE OF AID FOR MODELLING	286
5.1	Following a predictor through the analysis	292
6.	SAMPLE CONSIDERATIONS	295
6.1	Sample size	295
6.2	Sampling design	297
6.3	Outliers and robustness	297
7.	POSSIBLE DEVELOPMENTS IN AID	298
	REFERENCES	299

CHAPTER 6 CLUSTER ANALYSIS 301

Douglas M Hawkins, Michael W Muller, J Adri ten Krooden

1.	INTRODUCTION	301
2.	SPECIFIC CLUSTERING METHODS	306
2.1	Hierarchical methods	306
2.1.0	Definition of hierarchical methods	306
2.1.1	Single linkage cluster analysis	307
2.1.2	Mode analysis	309
2.1.3	Group average linkage	311
2.1.4	Median linkage	312
2.1.5	Ward's method	313
2.1.6	Lance and Williams' flexible method	315
2.2	Non-hierarchical methods	316
2.2.0	Non-hierarchical methods in general	316
2.2.1	Mixture analysis	316
2.2.2	Latent profile analysis	320
2.2.3	Allocation methods	322
2.2.4	Q-mode factor analysis	328

Cambridge University Press
978-0-521-09070-4 - Topics in Applied Multivariate Analysis
Edited by Douglas M. Hawkins
Frontmatter
[More information](#)

3. PRACTICAL CONSIDERATIONS	329
4. INFERENCE ON MODEL FIT AND NUMBER OF CLUSTERS	331
4.1 Tests of normal fit of profile data	332
4.2 Tests on number of clusters	334
4.3 Illustration	339
5. SEGMENTATION	342
5.1 Introduction	342
5.2 Segmentation of a sequence	343
5.3 Methods for segmentation	345
5.3.1 Global optimization	345
5.3.2 Split moving window	345
5.3.3 Hierarchic methods	346
5.4 Segmentation of spatial data	347
5.4.1 Methods for segmentation	347
5.5 Illustrations	348
6. FURTHER READING	351
REFERENCES	351
INDEX	355