

CONTENTS

PREFACE		xi
ACKNOWLEDGMENTS		xiii
Chapter 1.	The Broad View of Cluster Analysis	
1.1	Category Sorting Problems	1
1.2	Need for Cluster Analysis Algorithms	3
1.3	Uses of Cluster Analysis	4
1.4	Literature of Cluster Analysis	6
1.5	Purpose of This Book	8
Chapter 2.	Conceptual Problems in Cluster Analysis	
2.1	Elements of a Cluster Analysis	10
2.2	Illustrative Example	16
2.3	Some Philosophical Observations	18
2.4	A Note on Optimality and Intuition	24
Chapter 3.	Variables and Scales	
3.1	Classification of Variables	26
3.2	Scale Conversions	30
3.3	The Application of Scale Conversions	68
Chapter 4.	Measures of Association among Variables	
4.1	Measures between Ratio and Interval Variables	71
4.2	Measures between Nominal Variables	75
4.3	Measures between Binary Variables	83
4.4	Strategies for Mixed Variable Data Sets	92

Chapter 5.	Measures of Association among Data Units	
5.1	Metric Measures for Interval Variables	99
5.2	Nonmetric Measures for Interval Variables	110
5.3	Measures Using Binary Variables	114
5.4	Measures Using Nominal Variables	122
5.5	Mixed Variable Strategies	127
Chapter 6.	Hierarchical Clustering Methods	
6.1	The Central Agglomerative Procedure	132
6.2	The Stored Matrix Approach	134
6.3	The Stored Data Approach	145
6.4	The Sorted Matrix Approach	149
6.5	Other Approaches	152
Chapter 7.	Nonhierarchical Clustering Methods	
7.1	Initial Configurations	157
7.2	Nearest Centroid Sorting—Fixed Number of Clusters	160
7.3	Nearest Centroid Sorting—Variable Number of Clusters	167
7.4	Other Approaches to Nonhierarchical Clustering	173
Chapter 8.	Promoting Interpretation of Clustering Results	
8.1	Aids to Interpreting Hierarchical Classifications	177
8.2	An Aid to Interpreting a Partition of Data Units into Clusters	180
Chapter 9.	Strategies for Using Cluster Analysis	
9.1	Sequential Clustering of Data Units	182
9.2	Complementary Use of Several Clustering Methods	187
9.3	Cluster Analysis as an Adjunct to Other Statistical Methods	190
9.4	Clustering with Respect to an External Criterion	194
9.5	The Need for Research on Strategies	198
Chapter 10.	Comparative Evaluation of Cluster Analysis Methods	
10.1	An Approach to the Evaluation of Clustering Methods	200
10.2	Quantitative Assessment of Performance for Clustering Methods	202
10.3	List of Candidate Characteristic for Problems and Methods	209
10.4	The Evaluation Task Lying Ahead	213

Appendix A. Correlation and Nominal Variables

A.1	The Fundamental Analysis	216
A.2	The Problem of Isolated Cells	221
A.3	Deflating the Squared Correlation	226

Appendix B. Programs for Scale Conversions

B.1	Partitions of the Truncated Normal Distribution	228
B.2	Iterative Improvement of a Partition	229
	Program CUTS	230
	Function ERF	230
	Program DIVIDE	231
	Subroutine TEST	233
	Subroutine SORT	234
	Function PSUMSQ	234

Appendix C. Programs for Association Measures among Nominal and Interval Variables

C.1	General Design Features	235
C.2	Deck Setup and Utilization	236
	Subroutine GCORR	237
	Subroutine INPTR	241
	Subroutine NCAT	242
	Subroutine EIGEN	242
	Subroutine VSORT	245
	Function CORXX	246
	Function CORKX	247
	Function CORKK	248

Appendix D. Programs for Association Measures Involving Binary Variables

D.1	Bit-Level Storage	259
D.2	Computing Association Measures	261
D.3	Use of the Program	262
	Program BINARY	262
	Subroutine BDATA	264
	Function Subprogram KOUNT	267
	Function BASSN	269

Appendix E.	Programs for Hierarchical Cluster Analysis	
E.1	Stored Similarity Matrix Approach	276
E.2	Stored Data Approach	276
E.3	Sorted Matrix Approach	277
	Subroutine CNTRL	278
	Subroutine CLSTR	281
	Function LFIND	283
	Subroutine METHOD	283
	Subroutine MANAGE	290
	Subroutine GROUP	293
	Subroutine PROC	295
	Subroutine ALLIN1	300
	Subroutine PREP	303
Appendix F.	Programs for Nonhierarchical Clustering	
	Subroutine EXEC	307
	Subroutine RESULT	310
	Subroutine KMEAN	311
Appendix G.	Programs to Aid Interpretation of Clustering Results	
G.1	A Program for Manipulating Hierarchical Trees	326
G.2	Permuting the Similarity Matrix	329
G.3	Error Sum of Squares Analysis	330
G.4	Analysis of a Given Partition	330
	Subroutine DETAIL	330
	Subroutine READCM	333
	Subroutine TREE	333
	Program PERMUTE	337
	Subroutine MIXIN	339
	Function LFIND	340
	Program ERROR	340
	Program POSTDU	342
Appendix H.	Relations Among Cluster Analysis Programs	346
References		347
INDEX		355