

WILEY SERIES IN COMPUTATIONAL STATISTICS



Lynne Billard and Edwin Diday

SYMBOLIC DATA ANALYSIS

CONCEPTUAL STATISTICS AND DATA MINING

 WILEY

Contents

1	Introduction	1
	References	6
2	Symbolic Data	7
2.1	Symbolic and Classical Data	8
2.1.1	Types of data	8
2.1.2	Dependencies in the data	30
2.2	Categories, Concepts, and Symbolic Objects	34
2.2.1	Preliminaries	34
2.2.2	Descriptions, assertions, extents	35
2.2.3	Concepts of concepts	45
2.2.4	Some philosophical aspects	50
2.2.5	Fuzzy, imprecise, and conjunctive data	53
2.3	Comparison of Symbolic and Classical Analyses	56
	Exercises	66
	References	67
3	Basic Descriptive Statistics: One Variate	69
3.1	Some Preliminaries	69
3.2	Multi-Valued Variables	73
3.3	Interval-Valued Variables	77
3.4	Modal Multi-Valued Variables	92
3.5	Modal Interval-Valued Variables	96
	Exercises	105
	References	106
4	Descriptive Statistics: Two or More Variates	107
4.1	Multi-Valued Variables	107
4.2	Interval-Valued Variables	109
4.3	Modal Multi-Valued Variables	113

4.4	Modal Interval-Valued Variables	116
4.5	Baseball Interval-Valued Dataset	123
4.5.1	The data: actual and virtual datasets	123
4.5.2	Joint histograms	127
4.5.3	Guiding principles	130
4.6	Measures of Dependence	131
4.6.1	Moment dependence	131
4.6.2	Spearman's rho and copulas	138
	Exercises	143
	References	144
5	Principal Component Analysis	145
5.1	Vertices Method	145
5.2	Centers Method	172
5.3	Comparison of the Methods	180
	Exercises	185
	References	186
6	Regression Analysis	189
6.1	Classical Multiple Regression Model	189
6.2	Multi-Valued Variables	192
6.2.1	Single dependent variable	192
6.2.2	Multi-valued dependent variable	195
6.3	Interval-Valued Variables	198
6.4	Histogram-Valued Variables	202
6.5	Taxonomy Variables	204
6.6	Hierarchical Variables	214
	Exercises	227
	References	229
7	Cluster Analysis	231
7.1	Dissimilarity and Distance Measures	231
7.1.1	Basic definitions	231
7.1.2	Multi-valued variables	237
7.1.3	Interval-valued variables	241
7.1.4	Mixed-valued variables	248
7.2	Clustering Structures	249
7.2.1	Types of clusters: definitions	249
7.2.2	Construction of clusters: building algorithms	256
7.3	Partitions	257
7.4	Hierarchy-Divisive Clustering	259
7.4.1	Some basics	259
7.4.2	Multi-valued variables	262
7.4.3	Interval-valued variables	265

CONTENTS

vii

7.5	Hierarchy–Pyramid Clusters.....	285
7.5.1	Some basics	285
7.5.2	Comparison of hierarchy and pyramid structures	290
7.5.3	Construction of pyramids	292
	Exercises	305
	References	306
Data Index		309
Author Index		311
Subject Index		313