1. Record Nr. UNINA9910144714603321

Autore	Billard L (Lynne), <1943->
Titolo	Symbolic data analysis [[electronic resource]] : conceptual statistics and data mining / / Lynne Billard, Edwin Diday
Pubbl/distr/stampa	Chichester, England ; ; Hoboken, NJ, : John Wiley & Sons Inc., c2006
ISBN	1-280-59231-1 9786613622143 0-470-09018-9 0-470-09017-0
Descrizione fisica	1 online resource (331 p.)
Collana	Wiley Series in Computational Statistics ; ; v.654 Wiley series in computational statistics
Altri autori (Persone)	DidayE
Disciplina	519.5/35 519.535
Soggetti	Data mining Multivariate analysis Electronic books.
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Note generali	Description based upon print version of record.
Nota di bibliografia	Includes bibliographical references and indexes.
Nota di contenuto	Symbolic Data Analysis; Contents; 1 Introduction; References; 2 Symbolic Data; 2.1 Symbolic and Classical Data; 2.1.1 Types of data; 2.1.2 Dependencies in the data; 2.2 Categories, Concepts, and Symbolic Objects; 2.2.1 Preliminaries; 2.2.2 Descriptions, assertions, extents; 2.2.3 Concepts of concepts; 2.2.4 Some philosophical aspects; 2.2.5 Fuzzy, imprecise, and conjunctive data; 2.3 Comparison of Symbolic and Classical Analyses; Exercises; References; 3 Basic Descriptive Statistics: One Variate; 3.1 Some Preliminaries; 3.2 Multi-Valued Variables; 3.3 Interval-Valued Variables 3.4 Modal Multi-Valued Variables3.5 Modal Interval-Valued Variables; Exercises; References; 4 Descriptive Statistics: Two or More Variates; 4.1 Multi-Valued Variables; 4.2 Interval-Valued Variables; 4.3 Modal Multi-Valued Variables; 4.4 Modal Interval-Valued Variables; 4.5 Baseball Interval-Valued Dataset; 4.5.1 The data: actual and virtual datasets; 4.5.2 Joint histograms; 4.5.3 Guiding principles; 4.6 Measures of Dependence; 4.6.1 Moment dependence; 4.6.2 Spearman's rho and

copulas; Exercises; References; 5 Principal Component Analysis; 5.1 Vertices Method; 5.2 Centers Method
5.3 Comparison of the MethodsExercises; References; 6 Regression Analysis; 6.1 Classical Multiple Regression Model; 6.2 Multi-Valued Variables; 6.2.1 Single dependent variable; 6.2.2 Multi-valued dependent variable; 6.3 Interval-Valued Variables; 6.4 Histogram-Valued Variables; 6.5 Taxonomy Variables; 6.6 Hierarchical Variables; Exercises; References; 7 Cluster Analysis; 7.1 Dissimilarity and Distance Measures; 7.1.1 Basic definitions; 7.1.2 Multi-valued variables; 7.1.3 Interval-valued variables; 7.1.4 Mixed-valued variables; 7.2 Clustering Structures; 7.2.1 Types of clusters: definitions
7.2.2 Construction of clusters: building algorithms7.3 Partitions; 7.4 Hierarchy-Divisive Clustering; 7.4.1 Some basics; 7.4.2 Multi-valued variables; 7.4.3 Interval-valued variables; 7.5 Hierarchy-Pyramid Clusters; 7.5.1 Some basics; 7.5.2 Comparison of hierarchy and pyramid structures; 7.5.3 Construction of pyramids; Exercises; References; Data Index; Author Index; Subject Index

Sommario/riassunto

With the advent of computers, very large datasets have become routine. Standard statistical methods don't have the power or flexibility to analyse these efficiently, and extract the required knowledge. An alternative approach is to summarize a large dataset in such a way that the resulting summary dataset is of a manageable size and yet retains as much of the knowledge in the original dataset as possible. One consequence of this is that the data may no longer be formatted as single values, but be represented by lists, intervals, distributions, etc. The summarized data have their own internal s