

1. Record Nr.	UNINA9910132172203321
Autore	Myatt Glenn J. <1969->
Titolo	Making sense of data I : a practical guide to exploratory data analysis and data mining / / Glenn J. Myatt, Wayne P. Johnson
Pubbl/distr/stampa	Hoboken, New Jersey : , : Wiley, , 2014 ©2014
ISBN	1-118-42200-7 1-118-42201-5
Edizione	[Second edition.]
Descrizione fisica	1 online resource (250 p.)
Disciplina	006.3/12
Soggetti	Data mining Mathematical statistics
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 index.
Nota di contenuto	Making Sense of Data I; Contents; Preface; 1 Introduction; 1.1 Overview; 1.2 Sources of Data; 1.3 Process for Making Sense of Data; 1.3.1 Overview; 1.3.2 Problem Definition and Planning; 1.3.3 Data Preparation; 1.3.4 Analysis; 1.3.5 Deployment; 1.4 OVERVIEW OF BOOK; 1.4.1 Describing Data; 1.4.2 Preparing Data Tables; 1.4.3 Understanding Relationships; 1.4.4 Understanding Groups; 1.4.5 Building Models; 1.4.6 Exercises; 1.4.7 Tutorials; 1.5 Summary; Further Reading; Exercises; Exercises; Exercises; Exercises; 2 Describing Data; 2.1 Overview; 2.2 Observations and Variables 2.3 Types of Variables2.4 Central Tendency; 2.4.1 Overview; 2.4.2 Mode; 2.4.3 Median; 2.4.4 Mean; 2.5 Distribution of the Data; 2.5.1 Overview; 2.5.2 Bar Charts and Frequency Histograms; 2.5.3 Range; 2.5.4 Quartiles; 2.5.5 Box Plots; 2.5.6 Variance; 2.5.7 Standard Deviation; 2.5.8 Shape; 2.6 Confidence Intervals; 2.7 Hypothesis Tests; Further Reading; Further Reading; Further Reading; Further Reading; 3 Preparing Data Tables; 3.1 Overview; 3.2 Cleaning the Data; 3.3 Removing Observations and Variables; 3.4 Generating Consistent Scales Across Variables; 3.5 New Frequency Distribution 3.6 Converting Text to Numbers3.7 Converting Continuous Data to Categories; 3.8 Combining Variables; 3.9 Generating Groups; 3.10 Preparing Unstructured Data; 4 Understanding Relationships; 4.1

Overview; 4.2 Visualizing Relationships Between Variables; 4.2.1 Scatterplots; 4.2.2 Summary Tables and Charts; 4.2.3 Cross-Classification Tables; 4.3 Calculating Metrics About Relationships; 4.3.1 Overview; 4.3.2 Correlation Coefficients; 4.3.3 Kendall Tau; 4.3.4 t-Tests Comparing Two Groups; 4.3.5 ANOVA; 4.3.6 Chi-Square; 5 Identifying and Understanding Groups; 5.1 Overview; 5.2 Clustering 5.2.1 Overview 5.2.2 Distances; 5.2.3 Agglomerative Hierarchical Clustering; 5.2.4 k-Means Clustering; 5.3 Association Rules; 5.3.1 Overview; 5.3.2 Grouping by Combinations of Values; 5.3.3 Extracting and Assessing Rules; 5.3.4 Example; 5.4 Learning Decision Trees from Data; 5.4.1 Overview; 5.4.2 Splitting; 5.4.3 Splitting Criteria; 5.4.4 Example; Exercises; Further Reading; 6 Building Models from Data; 6.1 Overview; 6.2 Linear Regression; 6.2.1 Overview; 6.2.2 Fitting a Simple Linear Regression Model; 6.2.3 Fitting a Multiple Linear Regression Model; 6.2.4 Assessing the Model Fit 6.2.5 Testing Assumptions 6.2.6 Selecting and Assessing Independent Variables; 6.3 Logistic Regression; 6.3.1 Overview; 6.3.2 Fitting a Simple Logistic Regression Model; 6.3.3 Fitting and Interpreting Multiple Logistic Regression Models; 6.3.4 Significance of Model and Coefficients; 6.3.5 Classification; 6.4 k-Nearest Neighbors; 6.4.1 Overview; 6.4.2 Training; 6.4.3 Predicting; 6.5 Classification and Regression Trees; 6.5.1 Overview; 6.5.2 Predicting; 6.5.3 Example; 6.6 Other Approaches; 6.6.1 Neural Networks; 6.6.2 Support Vector Machines; 6.6.3 Discriminant Analysis; 6.6.4 Naive Bayes 6.6.5 Random Forests

Sommario/riassunto

With a focus on the needs of educators and students, Making Sense of Data presents the steps and issues that need to be considered in order to successfully complete a data analysis or data mining project. This Second Edition focuses on basic data analysis approaches that are necessary to complete a diverse range of projects. New examples have been added to illustrate the different approaches, and there is considerably more emphasis on hands-on software tutorials to provide real-world exercises. Via the related Web site, the book is accompanied by Traceis software, data sets, a
