1. Record Nr. UNINA9910139454503321

Autore	Huber Peter J
Titolo	Data analysis [[electronic resource]] : what can be learned from the past 50 years / / Peter J. Huber
Pubbl/distr/stampa	Hoboken, New Jersey : , : Wiley, , c2011
ISBN	1-283-10931-X 9786613109316 1-118-01825-7 1-118-01824-9
Edizione	[First edition]
Descrizione fisica	1 online resource (235 pages)
Collana	Wiley series in probability and statistics.
Disciplina	519.5 519.509
Soggetti	Mathematical statistics - History Mathematical statistics - Philosophy Numerical analysis - Methodology
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	DATA ANALYSIS: What Can Be Learned From the Past 50 Years; CONTENTS; Preface; 1 What is Data Analysis?; 1.1 Tukey's 1962 paper; 1.2 The Path of Statistics; 2 Strategy Issues in Data Analysis; 2.1 Strategy in Data Analysis; 2.2 Philosophical issues; 2.2.1 On the theory of data analysis and its teaching; 2.2.2 Science and data analysis; 2.2.3 Economy of forces; 2.3 Issues of size; 2.4 Strategic planning; 2.4.1 Planning the data collection; 2.4.2 Choice of data and methods; 2.4.3 Systematic and random errors; 2.4.4 Strategic reserves; 2.4.5 Human factors; 2.5 The stages of data analysis 2.5.1 Inspection2.5.2 Error checking; 2.5.3 Modification; 2.5.4 Comparison; 2.5.5 Modeling and Model fitting; 2.5.6 Simulation; 2.5.7 What-if analyses; 2.5.8 Interpretation; 2.5.9 Presentation of conclusions; 2.6 Tools required for strategy reasons; 2.6.1 Ad hoc programming; 2.6.2 Graphics; 2.6.3 Record keeping; 2.6.4 Creating and keeping order; 3 Massive Data Sets; 3.1 Introduction; 3.2 Disclosure: Personal experiences; 3.3 What is massive? A classification of size; 3.4 Obstacles to scaling; 3.4.1 Human limitations: visualization;

3.4.2 Human - machine interactions

3.4.3 Storage requirements3.4.4 Computational complexity; 3.4.5 Conclusions; 3.5 On the structure of large data sets; 3.5.1 Types of data; 3.5.2 How do data sets grow?; 3.5.3 On data organization; 3.5.4 Derived data sets; 3.6 Data base management and related issues; 3.6.1 Data archiving; 3.7 The stages of a data analysis; 3.7.1 Planning the data collection; 3.7.2 Actual collection; 3.7.3 Data access; 3.7.4 Initial data checking; 3.7.5 Data analysis proper; 3.7.6 The final product: presentation of arguments and conclusions; 3.8 Examples and some thoughts on strategy; 3.9 Volume reduction

3.10 Supercomputers and software challenges3.10.1 When do we need a Concorde?; 3.10.2 General Purpose Data Analysis and Supercomputers; 3.10.3 Languages, Programming Environments and Databased Prototyping; 3.11 Summary of conclusions; 4 Languages for Data Analysis; 4.1 Goals and purposes; 4.2 Natural languages and computing languages; 4.2.1 Natural languages; 4.2.2 Batch languages; 4.2.3 Immediate languages; 4.2.4 Language and literature; 4.2.5 Object orientation and related structural issues; 4.2.6 Extremism and compromises, slogans and reality; 4.2.7 Some conclusions; 4.3 Interface issues

4.3.1 The command line interface4.3.2 The menu interface; 4.3.3 The batch interface and programming environments; 4.3.4 Some personal experiences; 4.4 Miscellaneous issues; 4.4.1 On building blocks; 4.4.2 On the scope of names; 4.4.3 On notation; 4.4.4 Book-keeping problems; 4.5 Requirements for a general purpose immediate language; 5 Approximate Models; 5.1 Models; 5.2 Bayesian modeling; 5.3 Mathematical statistics and approximate models; 5.4 Statistical significance and physical relevance; 5.5 Judicious use of a wrong model; 5.6 Composite models; 5.7 Modeling the length of day

5.8 The role of simulation

Sommario/riassunto

This book explores the many provocative questions concerning the fundamentals of data analysis. It is based on the time-tested experience of one of the gurus of the subject matter. Why should one study data analysis? How should it be taught? What techniques work best, and for whom? How valid are the results? How much data should be tested? Which machine languages should be used, if used at all? Emphasis on apprenticeship (through hands-on case studies) and anecdotes (through real-life applications) are the tools that Peter J. Huber uses in this volume. Concern with specific statistical techniq