1. Record Nr. UNINA9910271040103321

Autore	Loo Mark van der <1976->
Titolo	Statistical data cleaning with applications in R / / Mark van der Loo, Edwin de Jonge
Pubbl/distr/stampa	Hoboken, New Jersey : , : Wiley, , 2018 ©2018
ISBN	1-118-89713-7 1-118-89714-5 1-118-89712-9
Edizione	[1st edition]
Descrizione fisica	1 online resource (301 pages) : illustrations
Disciplina	519.50285/5133
Soggetti	Statistics - Data processing R (Computer program language) Electronic books.
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di bibliografia	Includes bibliographical references and index.
Sommario/riassunto	A comprehensive guide to automated statistical data cleaning The production of clean data is a complex and time-consuming process that requires both technical know-how and statistical expertise. Statistical Data Cleaning brings together a wide range of techniques for cleaning textual, numeric or categorical data. This book examines technical data cleaning methods relating to data representation and data structure. A prominent role is given to statistical data validation, data cleaning based on predefined restrictions, and data cleaning strategy. Key features: Focuses on the automation of data cleaning methods, including both theory and applications written in R. Enables the reader to design data cleaning processes for either one-off analytical purposes or for setting up production systems that clean data on a regular basis. Explores statistical techniques for solving issues such as incompleteness, contradictions and outliers, integration of data cleaning components and quality monitoring. Supported by an accompanying website featuring data and R code. This book enables data scientists and statistical analysts working with data to deepen

their understanding of data cleaning as well as to upgrade their practical data cleaning skills. It can also be used as material for a course in data cleaning and analyses.