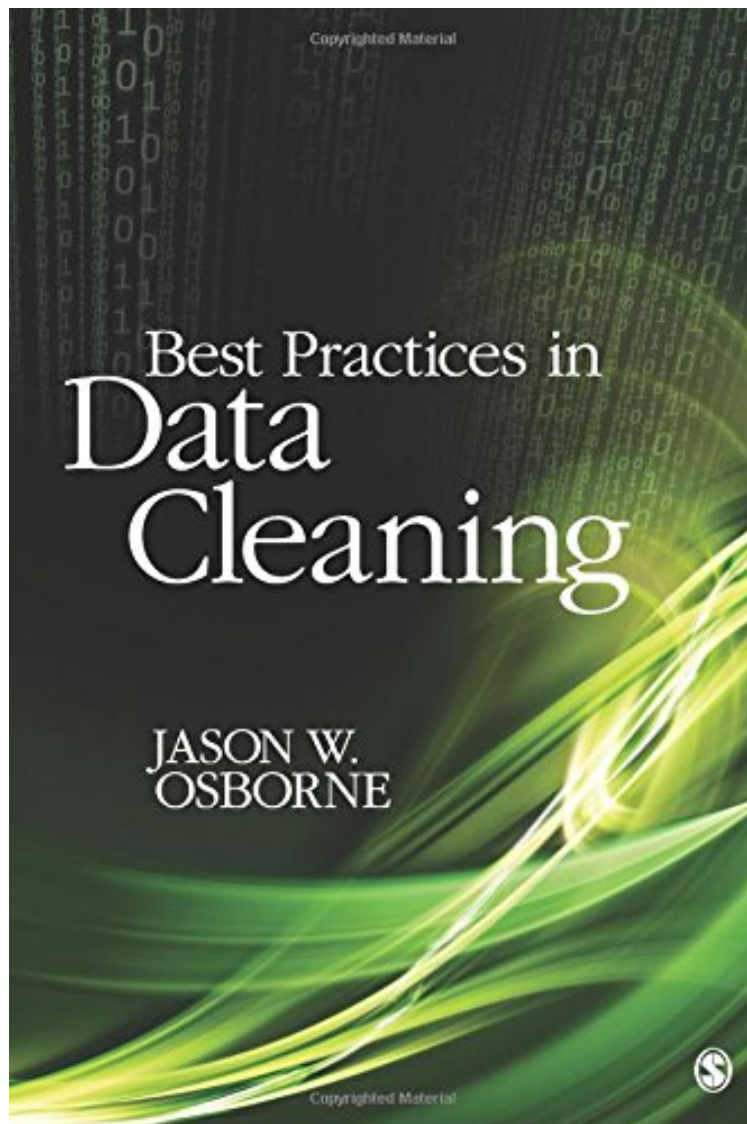


(Pdf free) Best Practices in Data Cleaning: A Complete Guide to Everything You Need to Do Before and After Collecting Your Data

Best Practices in Data Cleaning: A Complete Guide to Everything You Need to Do Before and After Collecting Your Data

Jason W. Osborne

DOC | *audiobook | ebooks | Download PDF | ePub



#405424 in Books SAGE Publications, Inc 2012-01-10 2012-01-10Original language:EnglishPDF # 1 9.00 x .67 x 6.00l, .75 #File Name: 1412988012296 pages | File size: 29.Mb

Jason W. Osborne : Best Practices in Data Cleaning: A Complete Guide to Everything You Need to Do Before and After Collecting Your Data before purchasing it in order to gage whether or not it would be worth my time, and all praised Best Practices in Data Cleaning: A Complete Guide to Everything You Need to Do Before and After Collecting Your Data:

5 of 8 people found the following review helpful. Comprehensive and logical
By Michael McKean
Best Practices in Data Cleaning by Jason Osborne provides a comprehensive guide to data cleaning. Although I have had a great deal of training associated with the process of setting up and reviewing data collection and analysis I had been away from the field for several years, and recent work required that I consult prior to beginning a new project. This book made the process of getting back up to speed enjoyable. I was very pleased by the step-by-step explanations provided, and found the book to be an excellent resource.
Mike M. Eng. (Engineering Management), M.Sc. (Computer Science), MSW (Candidate)
15 of 15 people found the following review helpful. Worthwhile book with a misleading title
By Dimitri Shvorob
Well, it cost me \$4 in postage to discover that "Best practices in data cleaning" is not about data cleaning, but mostly about the challenges of working with survey data. The author acknowledges upfront that his "data cleaning" includes activities undertaken both before and after data collection, and his first topic, power and sample size analysis, drives the point home. Among the other topics, outlier detection alone is what I would call data cleaning; handling of missing data and normalizing transformations - maybe. Despite the middling rating, I like the book. Given the abundant evidence of lacking statistical expertise among social-sciences researchers - take the references cited here, or in "Understanding the new statistics" by Geoff Cumming - any educational effort is welcome, especially when delivered with an accessible, friendly writing style. The misgivings come from the titular misnomer - SAGE, I want my \$4 back - and, secondly, dissatisfaction with the statistics on display. I am not talking about the introductory nature of discussion - meaning that you will need to consult other references to understand how a particular method, and how it can be implemented - but about the suspicion that presented repertoire is a bit behind the curve. Consider "Data analysis using regression and multilevel/hierarchical models" by Gelman and Hill, for example - it's hard to believe that "Best practices in data cleaning" is more recent. R, simulation-based methods, "robust" or non-parametric methods, "exact" tests - absent or mentioned in a few words. ("Winsorizing", an antique "robust" method, is there, misspelled, and there is a simulation done in SPSS - filling four pages (!) with random numbers. It all feels very 1990s). Overall, I appreciate "Best practices in data cleaning" as an accessible "awareness-builder" offering practical advice and many useful references - while flagging some reservations about said advice, or rather about what is left unsaid. Applied researchers in psychometrics should take a look. Page visitors looking for a book about data cleaning - check out "Data quality assessment" by Arkady Maydanchik. PS. Normalizing transformations are given a lot of space in the book, and I wonder if the discussion is entirely accurate. On page 90, for example, the author seems to suggest that normality of data is a requirement of conventional statistical tests. I'd say that things are more nuanced - the central-limit theorem gets one asymptotic normality, but one deals with finite, and often small, samples - and note that since one is not always going to work in univariate setting (e.g, drawing histograms instead of running regressions), it's helpful to point out the difference between conditional and unconditional normality.

Many researchers jump from data collection directly into testing hypothesis without realizing these tests can go profoundly wrong without clean data. This book provides a clear, accessible, step-by-step process of important best practices in preparing for data collection, testing assumptions, and examining and cleaning data in order to decrease error rates and increase both the power and replicability of results. Jason W. Osborne, author of the handbook Best Practices in Quantitative Methods (SAGE, 2008) provides easily-implemented suggestions that are evidence-based and will motivate change in practice by empirically demonstrating for each topic the benefits of following best practices and the potential consequences of not following these guidelines.

This book provides the perfect bridge between the formal study of statistics and the practice of statistics. It fills the gap left by many of the traditional texts that focus either on the technical presentation or recipe-driven presentation of topics. --Elizabeth M. Flow-Delwiche (10/27/2011)"The first comprehensive and generally accessible text in this area. --J. Michael Hardin (10/27/2011)" This book provides the perfect bridge between the formal study of statistics and the practice of statistics. It fills the gap left by many of the traditional texts that focus either on the technical presentation or recipe-driven presentation of topics. --Elizabeth M. Flow-Delwiche (10/27/2011)" The first comprehensive and generally accessible text in this area. --J. Michael Hardin (10/27/2011)" "This book provides the perfect bridge between the formal study of statistics and the practice of statistics. It fills the gap left by many of the traditional texts that focus either on the technical presentation or recipe-driven presentation of topics."--Elizabeth M. Flow-Delwiche (10/27/2011)"The first comprehensive and generally accessible text in this area."--J. Michael Hardin (10/27/2011)-This book provides the perfect bridge between the formal study of statistics and the practice of statistics. It fills the gap left by many of the traditional texts that focus either on the technical presentation or recipe-driven presentation of topics.---Elizabeth M. Flow-Delwiche (10/27/2011)-The first comprehensive and generally accessible text in this area.---J. Michael Hardin (10/27/2011) This book provides the perfect bridge between the formal study of statistics and the practice of statistics. It fills the gap left by many of the traditional texts that focus either on the technical presentation or recipe-driven presentation of topics. (Elizabeth M. Flow-Delwiche 2011-10-27)The first comprehensive and generally accessible text in this area. (J. Michael Hardin 2011-10-27)About the Author Jason W. Osborne is Associate Provost and Dean of the Graduate School at Clemson University in Clemson, South Carolina. He

is also Professor of Applied Statistics in the Department of Mathematical Sciences, with a secondary appointment in Public Health Science. He teaches and publishes on "best practices" in quantitative and applied research methods. He has served as evaluator or consultant on projects in public education (K-12), instructional technology, higher education, nursing and health care, medicine and medical training, epidemiology, business and marketing, and jury selection in death penalty cases. He served as founding editor of *Frontiers in Quantitative Psychology and Measurement* and has been on the editorial boards of several other journals (such as *Practical Assessment, Research, and Evaluation*). Jason also publishes on identification with academics (how a student's self-concept impacts motivation to succeed in academics) and on issues related to social justice and diversity (such as *Stereotype Threat*). He is the very proud father of three, and holds the rank of third degree black belt in Songahm Tae Kwon Do. The rest is subject to change without notice (as Anne McCaffrey wrote in her bio).