

Cody's Data Cleaning Techniques Using SAS Software

By Ron Cody

ISBN 1-58025-600-7

Publication # 57198

Cost: \$34.95

This review is a composite analysis of two (Janet Stuelpner and Charles Patridge) people's thoughts on this book.

"Cody's Data Cleaning Techniques Using SAS Software" is an excellent tutorial for beginning through intermediate SAS programmers who have to deal with imperfect data which is quite common in many environments and applications. This is a very good step-by-step introduction to the world of data cleansing. Many advanced SAS professionals would also benefit from reading this book for its various creative techniques, and refreshing those sometimes long forgotten "tricks" one may have used in the past.

Even though Ron Cody's book illustrates all of his techniques using a typical clinical trial data set (file), the test data can easily be understood by most readers of this book and easily transferable to any industry. The style of the book is very conversational and this is one of very few drawbacks. If you have ever attended a presentation by Ron Cody, as you read, you feel like you are sitting in the auditorium with him as he speaks, but without the vacation pictures. The book could be a bit less familiar. To those of us who know all of the people to whom he refers, this is not a problem. However, without knowing the personalities and skill level of the people he quotes, it can be disconcerting.

The biggest advantage to this book is Ron's approach to addressing a cleansing technique for a particular problem by starting with a simple solution and building upon its simplicity to a very robust SAS macro solution. The book begins each section with a very elementary example upon which is built a more complicated task. This is very advantageous, especially for the beginner. There are a great number of helpful hints and methods that can be used by one and all. One criticism is the last example in the book, although pertinent to cleaning the data, is overkill. This last example brings the book from one that is geared toward the novice to one that is understood only by the expert. A big plus for this book is that it points out several approaches to a given problem highlighting the advantages and disadvantages of each approach. And, Ron should be commended in the way he provides comments to the numbered statements in each program making it very easy to read and understand.

Having more than twenty years of experience dealing with "messy" data and developing our own data "scrubbing" routines, we recommend having this book in your SAS library of books. It should be required reading for all new SAS programmers as we all too often have to deal with imperfect data. The cost of this book at \$34.95 can easily be justified by being able to utilize many of the actual source code examples illustrated within this book in your own applications. Overall, we give this book a THUMBS up!

One big complaint in reading this book in the **Books by Users**™ program is not having the various programs and data files easily accessible in electronic form. It would have been very easy for SAS Institute Inc. to provide a downloadable ZIP file on their web site making it easy to access, execute and, actually run the various programs while reading Ron Cody's newest book. This feature should be considered mandatory in all future books unless a diskette is included. It would be much more effective if the data and programs were in electronic format so the audience could try out the programs and the ideas put forth in the book.

Sincerely,

Charles Patridge
PDPC, Ltd.
172 Monce Road
Burlington, CT 06013
Phone: 860-673-9278
Email:

Charles_S_Patridge@prodigy.net

Janet Stuelpner
326 Old Norwalk Road
New Canaan, CT 06840
Phone: 203-966-7520
Email: Jstuelpner@att.net