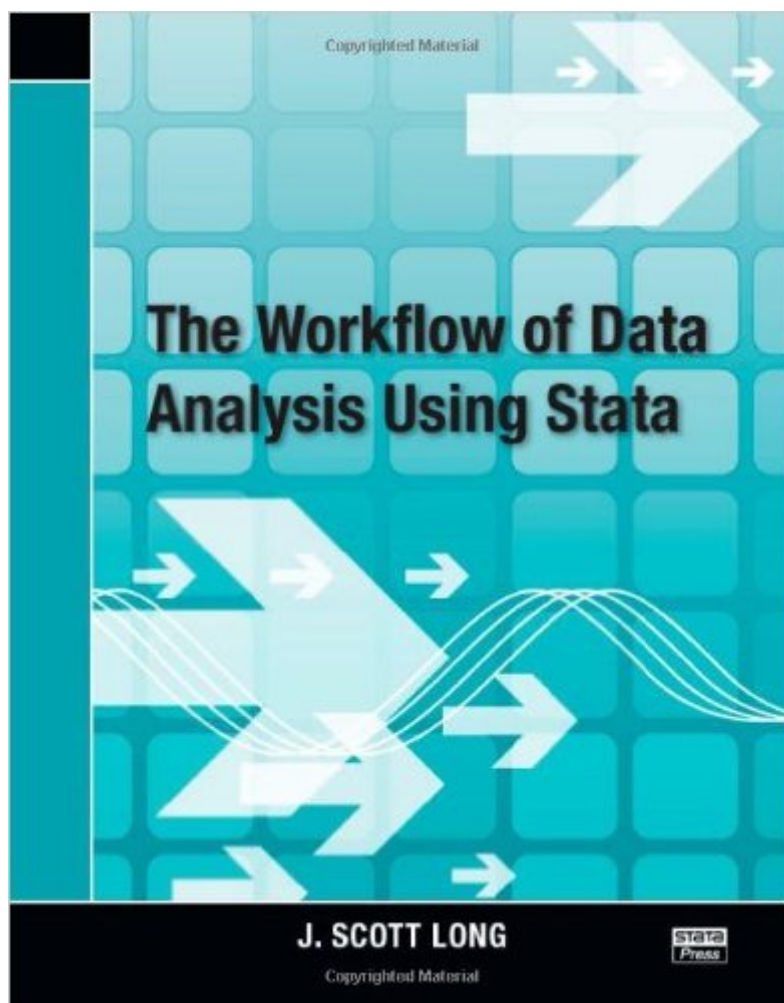# The Workflow Of Data Analysis Using Stata

Synopsis

The Workflow of Data Analysis Using Stata, by J. Scott Long, is an essential productivity tool for data analysts. Long presents lessons gained from his experience and demonstrates how to design and implement efficient workflows for both one-person projects and team projects. After introducing workflows and explaining how a better workflow can make it easier to work with data, Long describes planning, organizing, and documenting your work. He then introduces how to write and debug Stata do-files and how to use local and global macros. After a discussion of conventions that greatly simplify data analysis the author covers cleaning, analyzing, and protecting data.

Book Information

Paperback: 379 pages

Publisher: Stata Press; 1 edition (December 10, 2008)

Language: English

ISBN-10: 1597180475

ISBN-13: 978-1597180474

Product Dimensions: 7.2 x 0.9 x 9.2 inches

Shipping Weight: 1.8 pounds (View shipping rates and policies)

Average Customer Review: 4.8 out of 5 starsÂ Â See all reviewsÂ (19 customer reviews)

Best Sellers Rank: #238,991 in Books (See Top 100 in Books)   #139 inÂ Books > Computers & Technology > Software > Mathematical & Statistical   #606 inÂ Books > Textbooks > Science & Mathematics > Mathematics > Statistics   #906 inÂ Books > Science & Math > Mathematics > Applied > Probability & Statistics

Customer Reviews

Simply put, this book will drastically reduce the amount of time it will take you to write a do file. In fact, based on the suggestions that long has in this book I was able to write a do file that will run OLS and all the typical diagnostics automatically. All I have to do is assign the variable macros at the beginning of the do file. My summer project is to do the same thing for my other basic statistical models.The workflow of data analysis using Stata should be required for any course on regression or applied statistics that uses Stata. Any applied researcher using this program needs to buy this book.

This book is great. I highly recommend it for anyone dealing with large-scale collaborative data projects.The author covers everything from saving and filing your data (good life skills even for

those who don't use STATA) to documenting and cleaning datasets.The book reads fast and is interesting enough. Lots of general advice and then a few specific, detailed examples with full-fledged code that are easy to skip if they are not relevant (mercifully, he does not force you through the examples like some books do - they are fairly self-contained modules within the chapters). For each task, he tends to talk about the commands used (with relevant/most-helpful options), but then also about general rules for how to approach the task.The author's personal narrative dominates the book, but it is clear that he has a lot of experience and for the most part, it is helpful that the book is written that way. Some of the advice is shocking in its simplicity but its usefulness (i.e. "Never name a file 'Final'" lest you wind up with "Final - v1" "Final - v2" and "Final - really really final") - these are problems that are applicable to everyone. As such, although I have a reasonable amount of experience with STATA I felt that this book covered really useful basics in a way that was helpful but not tedious.Unsurprisingly for someone who writes about organization and workflow management, the book is also well-organized, and designed in such a way that you can read as you go, covering relevant chapters as you face the tasks within. I've been moving along at the pace of my project, highlighting commands as I read so that I can use it as a reference manual once I finish going through the first time. So far, the book has more than paid for itself in time and frustration saved.

Whether you use Stata or not, if you work with data and do research in the area of empirical social science (Economics, Political Science, Policy Studies) then this book is a must for you. And on top of that if you use Stata for your research, then the question is why you do not have this book yet in your library?This is probably one of of the few books that I ever bought which taught me something new in each page. It has definitely helped me to learn how to organize my work effectively and efficiently. It has also taught me a lot of Stata programming skills.

Short on statistics, and long on practical techniques, this thoughtful, incredibly well-organized guide should be required reading for anybody using STATA to complete a large project. My best guess is that for every hour you spend reading this book you will save 10 hours in your first analysis. Excellent on-line resources as well. Thank you Dr. Long!

All the examples are in Stata, but the workflow suggestions are useful for anyone who does statistical analysis and wants their analysis to be replicable. I have found the file management content particularly helpful. Be sure to check out the extras on Long's workflow for data analysis

website at Indiana University. I like the spreadsheet for planning directory structure that goes with chapter 2.

I found it extremely valuable for managing multiple projects and team members of multiple projects. In research, projects often continue while the team members may fluctuate. Following the workflow strategy has helped during times of transition. It also makes it easier to share people across multiple projects when there is consistency in organization. I've wondered how many hours this process has saved our team. We know where all our files are and which ones we can delete. That sounds so simple, right?

This book provide generalizable concept of good practice in data management. My project is a clinical research collecting data in Access and analyzing with STATA.I found in process of cleaning data- many files been created and became messy. The "Posting file" and "dual workflow (separate data management and statistical analysis" are really help me organize this project.

The book covers basic material in data management and organizing your work in STATA. Most of the material will prove beneficial to the novice STATA user or to someone who is about to start doing serious statistical analysis with STATA and wants to know a bit about organization and documentation of the work ahead. However, the material is also beneficial to the intermediate and advanced STATA user because it serves as a reminder of the right way of doing things when it comes to data and work organization that could potentially save tons of time in the future when looking for errors and trying to explain why things didnt turn out the way they should. On the down side, there is a lot of stuff in the book that can be found in other books and especially in the STATA Documentation series. Long tries hard to clarify each concept with many examples but in some places the examples go on forever. Also in some parts of the book the same material is covered more than once. Now this can be ok for the novice user but not for the expert. some of the material (for example why files get lost) could have stayed out of the book. I think that the same material could have been covered in 200 pages and not 350. Overall though, the book is a great resource for the novice researcher and can offer useful hints and tips about work organization to the more seasoned researchers.

The Workflow of Data Analysis Using Stata Data Analytics: Practical Data Analysis and Statistical Guide to Transform and Evolve Any Business Leveraging the Power of Data Analytics, Data

Science, ... (Hacking Freedom and Data Driven Book 2) Data Architecture: A Primer for the Data Scientist: Big Data, Data Warehouse and Data Vault An Introduction to Modern Econometrics Using Stata Big Data For Beginners: Understanding SMART Big Data, Data Mining & Data Analytics For improved Business Performance, Life Decisions & More! The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences Big Data, MapReduce, Hadoop, and Spark with Python: Master Big Data Analytics and Data Wrangling with MapReduce Fundamentals using Hadoop, Spark, and Python Statistics with STATA: Version 12 Python Data Analytics: Data Analysis and Science using pandas, matplotlib and the Python Programming Language Statistics for Ecologists Using R and Excel: Data Collection, Exploration, Analysis and Presentation (Data in the Wild) Microsoft Excel 2013 Data Analysis and Business Modeling: Data Analysis and Business Modeling (Introducing) Discovering Knowledge in Data: An Introduction to Data Mining (Wiley Series on Methods and Applications in Data Mining) LEARN IN A DAY! DATA WAREHOUSING. Top Links and Resources for Learning Data Warehousing ONLINE and OFFLINE: Use these FREE and PAID resources to Learn Data Warehousing in little to no time Data Just Right: Introduction to Large-Scale Data & Analytics (Addison-Wesley Data and Analytics) Introducing Data Science: Big Data, Machine Learning, and more, using Python tools Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython Excel Conditional Formatting: Tips You Can Use Immediately To Make Your Data Stand Out (Data Analysis With Excel Book 3) Just Plain Data Analysis: Finding, Presenting, and Interpreting Social Science Data Data Analysis Using Microsoft Excel: Updated for Office XP Adventures in Social Research: Data Analysis Using IBMÃ Â® SPSSÃ Â® Statistics