Brief contents

1 Introduction

2 Planning, organizing, and documenting

3 Writing and debugging do-files

4 Automating your work

5 Names, notes, and labels

6 Cleaning your data

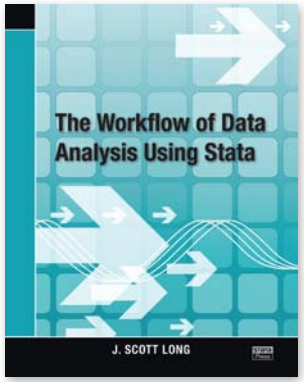
7 Analyzing data and presenting results

8 Protecting your files

9 Conclusions

A How Stata works

The Workflow of Data Analysis Using Stata



By J. Scott Long

Publisher: Stata Press
Copyright: 2009
ISBN-10: 1-59718-047-5
ISBN-13: 978-1-59718-047-4
Pages: 379; paperback
Price: **$52.00**

For more details or to order, visit us online at **www.stata-press.com/books/wdaus.html**.

STATA® Press

How to contact us

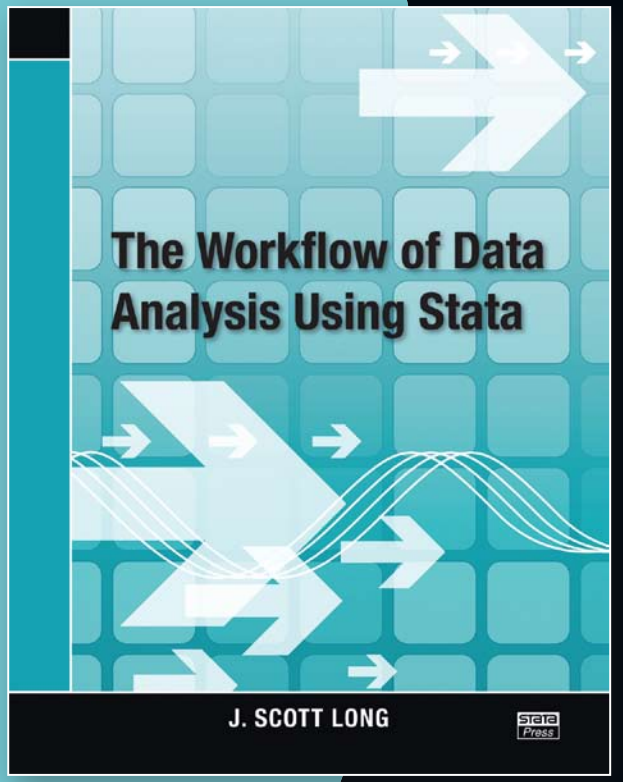
Stata Press
4905 Lakeway Drive
College Station, TX 77845
USA

800-STATAPC (800-782-8272, USA)
800-248-8272 (Canada)
979-696-4600 (Worldwide)

service@stata-press.com
www.stata-press.com

© Copyright 2009 StataCorp LP

Stata Press presents



About the author

J. Scott Long is Chancellor's Professor of Sociology and Statistics and Associate Vice Provost for Research at Indiana University–Bloomington. He has contributed articles to many journals, including *American Sociological Review*, *Social Forces*, *American Statistician*, and *Sociological Methods and Research*. He was editor of *Sociological Methods and Research* from 1987 to 1994. Dr. Long has authored or edited seven previous books on statistics, including the highly acclaimed *Regression Models for Categorical and Limited Dependent Variables*. In 2001, he received the Paul Lazarsfeld Memorial Award for Distinguished Contributions to Sociological Methodology. Each summer at the University of Michigan, he teaches workshops at the Inter-University Consortium for Political and Social Research Summer Program in Quantitative Methods of Social Research.

The book's audience

- Researchers who analyze data in social, behavioral, and biomedical sciences
- Students who analyze data in social, behavioral, and biomedical sciences
- Instructors who cover data management
  - ❖ Contains a unique introduction to organizing data-analytic projects
  - ❖ Is a great resource, with introductions to many Stata data-management commands and with several Stata examples
- Stata users who want to learn how to organize their data-analytic projects

Comment from the Stata technical group

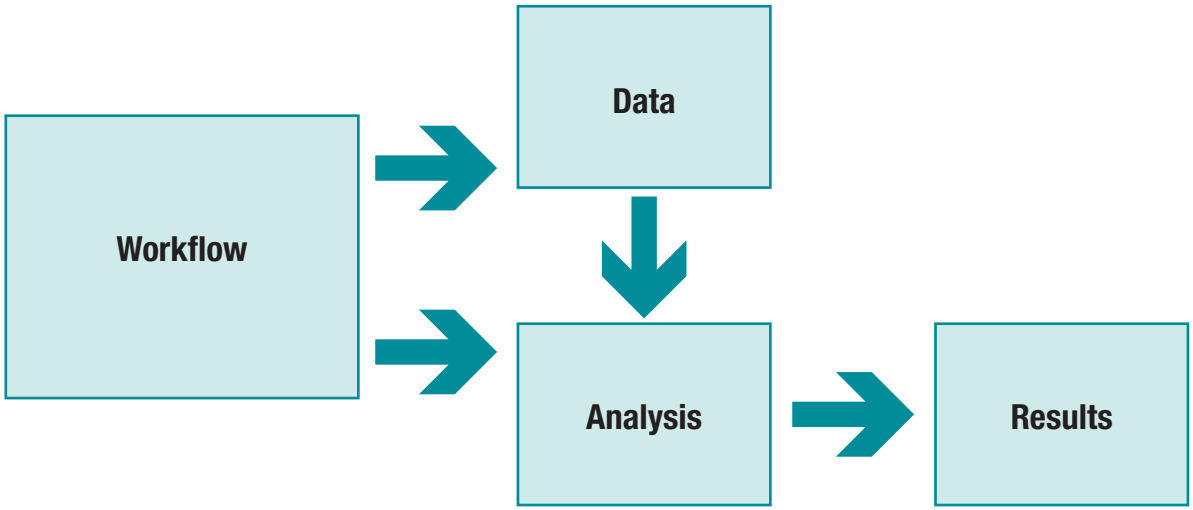
The Workflow of Data Analysis Using Stata, by J. Scott Long, is an essential productivity tool for data analysts. Aimed at anyone who analyzes data, this book presents an effective strategy for designing and doing data-analytic projects.

In this book, Long presents lessons gained from his experience with numerous academic publications, as a coauthor of the immensely popular *Regression Models for Categorical Dependent Variables Using Stata*, and as a coauthor of the SPOST routines, which are downloaded over 20,000 times a year.

A workflow of data analysis is a process for managing all aspects of data analysis. Planning, documenting, and organizing your work; cleaning the data; creating, renaming, and verifying variables; performing and presenting statistical analyses; producing replicable results; and archiving what you have done are all integral parts of your workflow.

Long shows how to design and implement efficient workflows for both one-person projects and team projects. Long guides you toward streamlining your workflow, because a good workflow is essential for replicating your work, and replication is essential for good science.

An efficient workflow reduces the time you spend doing data management and lets you produce datasets that are easier to analyze. When you methodically clean your data and carefully choose names and effective labels for your variables, the time you spend doing statistical and graphical analyses will be more productive and more enjoyable.

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. Long presents conventions that greatly simplify data analysis—conventions for naming, labeling, documenting, and verifying variables. He also covers cleaning, analyzing, and protecting your data.

While describing effective workflows, Long also introduces the concepts of basic data management using Stata and writing Stata do-files. Using real-world examples, Stata commands, and Stata scripts, Long illustrates effective techniques for managing your data and analyses. If you analyze data, this book is recommended for you.