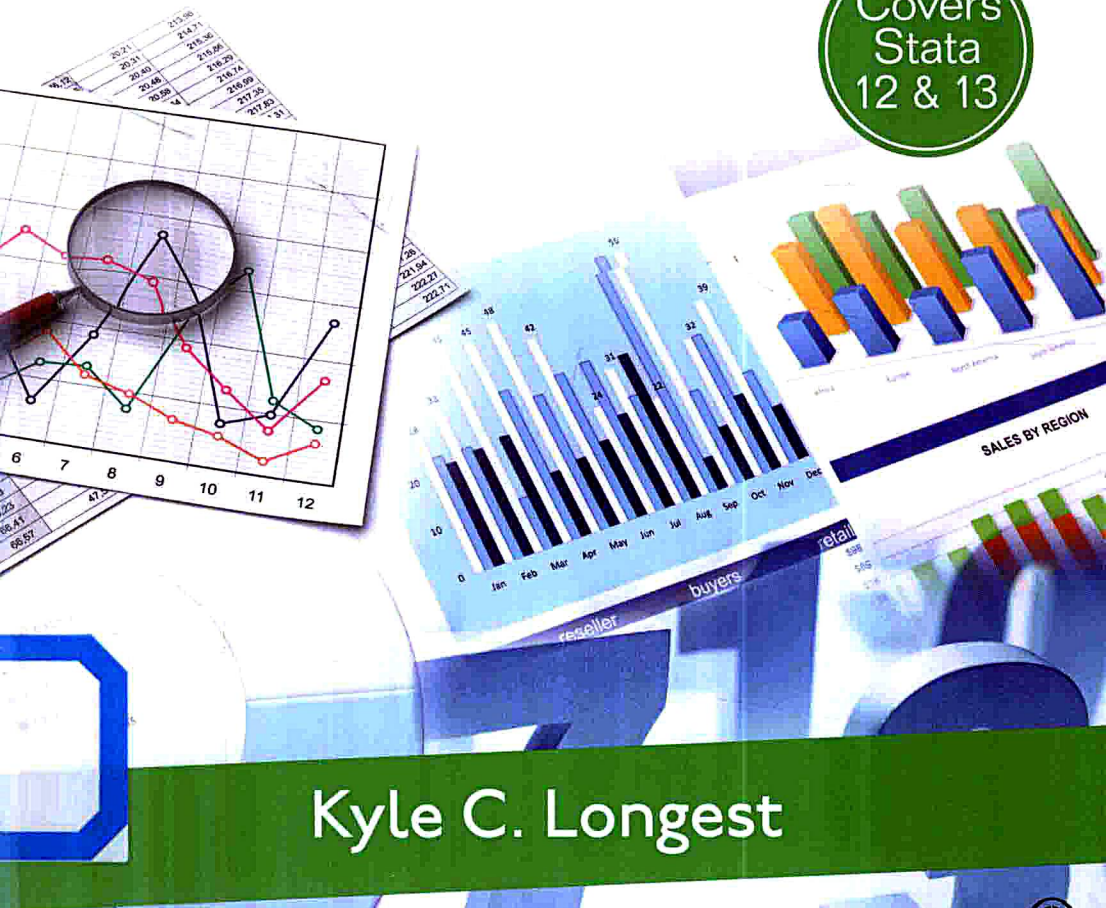
Second Edition

# Using Stata for Quantitative Analysis

Covers Stata 12 & 13




Kyle C. Longest

# Using Stata for Quantitative Analysis

**Kyle C. Longest**
*Furman University*

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC

FOR INFORMATION:

SAGE Publications, Inc.
2455 Teller Road
Thousand Oaks, California 91320
E-mail: order@sagepub.com

SAGE Publications Ltd.
1 Oliver's Yard
55 City Road
London EC1Y 1SP
United Kingdom

SAGE Publications India Pvt. Ltd.
B 1/I 1 Mohan Cooperative Industrial Area
Mathura Road, New Delhi 110 044
India

SAGE Publications Asia-Pacific Pte. Ltd.
33 Pekin Street #02-01
Far East Square
Singapore 048763



Printed in the United States of America

*Library of Congress Cataloging-in-Publication Data*

Longest, Kyle C., author.

Using Stata for quantitative analysis / Kyle C. Longest, Furman University.

pages cm
Includes bibliographical references and index.

ISBN 978-1-4833-5663-1 (pbk. : acid-free paper)

1. Stata. 2. Social sciences—Graphic methods—Computer programs. 3. Social sciences—Statistical methods—Computer programs. I. Title.

HA32.L66 2015
005.5'5—dc23 2014007519

This book is printed on acid-free paper.

Acquisitions Editor: Jerry Westby
Editorial Assistant: MaryAnn Vail
Production Editor: Olivia Weber-Stenis
Copy Editor: QuADS Prepress (P) Ltd.
Typesetter: C&M Digitals (P) Ltd.
Proofreader: Scott Oney
Cover Designer: Anupama Krishnan
Marketing Manager: Nicole Elliott


14 15 16 17 18 10 9 8 7 6 5 4 3 2 1

# Using Stata for Quantitative Analysis

# Detailed Contents

# *List of* A Closer Look *Boxes*

# Preface

## Motivation and Purpose

The motivation for this book, as I assume is true for most, came from a series of personal experiences. First, as a graduate student, I remember literally lying awake at night dreading the idea of using a computer program to conduct statistical analyses. The first statistics course I took required Stata to complete the assignments and the final research project. This necessity was so overwhelming at the time, in part, because there did not seem to be any straightforward, concise texts explaining the basics of Stata. During my time in graduate school, I came to be very familiar with Stata, even to the point that I developed a serious passion for both learning Stata and teaching it to students who were facing the same fears I once did. When I first began teaching a course on quantitative analysis, I was hoping to use Stata as a significant portion of the classroom experience and requirements. Yet in a somewhat mirrored experience from when I was a student, I soon realized that there still was not a manageable introductory text on the use of Stata for quantitative research.[1] Thus, I sought to contribute to filling this void by providing a straightforward, applied introduction to using Stata.

This book will be most beneficial to readers who are novices when it comes to Stata and are at least in the early stages of learning strategies for conducting quantitative analysis. It does assume that the reader has a working knowledge of basic statistical techniques and terminology. The organization and coverage of the book is guided by the content and ordering of topics found in most introductory social statistics textbooks. In this manner, it can serve as an excellent companion, for either a class or a self-learner, to such a textbook.

[1]Assuredly, there are several very good and effective texts on learning Stata. Virtually all of these, however, are aimed at experienced users or are so detailed and long that they are not helpful for a typical classroom in which teaching Stata is not the primary purpose.

To be clear, this book should not be used to learn statistics or quantitative analysis. Some basic assumptions and explanations are provided, but these should not be used in place of a more thorough coverage of each of the analytic strategies. The statistical grounding for this book is based primarily on Frankfort-Nachmias and Leon-Guerrero's *Social Statistics for a Diverse Society* (2010). The definitions and interpretations of the specific measures and tests are based on those presented in this text. Of course, any inaccuracies or mistakes are solely mine.

Also, this book does not attempt to cover every aspect of each Stata command that is introduced. More experienced users undoubtedly know shortcuts or alternative methods for the techniques that are presented. The given description has been geared to introduce complete novice users to Stata. This targeted audience requires that the explanation start with the basics before jumping into the advanced features. The presented commands and procedures are discussed because they are the most simplified strategies that effectively accomplish the pertinent goals.

## About the National Study of Youth and Religion

The data for this book come from the National Study of Youth and Religion (NSYR). The NSYR is a longitudinal, nationally representative telephone survey of U.S. young adults. There are three waves of data, all of which are publically available.

The variables that are used in the examples throughout this book come from the second follow-up survey of 2,532 young adults completed in the fall of 2007. At the time of this survey, the respondents were all between the ages of 18 and 24. Each respondent completed a computer-assisted telephone interviewing survey that lasted approximately an hour. This data set covers a broad array of topics, making it possible, across examples, to use variables pertinent to several disciplines. For example, it contains several standard self-esteem measures of interest to psychologists, a wide array of questions on religion useful for sociologists, numerous questions on finances (e.g., debt) applicable to economics, and measures of substance use behaviors that would be pertinent to social work or health researchers. The full data set and documentation can be downloaded from the Association of Religion Data Archives (http://www.thearda.com/Archive/Files/Descriptions/NSYRW3.asp).

The first wave of the survey sampled 3,290 U.S. English- and Spanish-speaking teenagers, ages 13 to 17. The sampling and survey were conducted from July 2002 to August 2003 using random-digit dialing, drawing on a sample of randomly generated telephone numbers representative of all noncellular phone numbers in the United States. The overall response rate of 57% for the

first survey is lower than desired, but it is similar to other current nationally based surveys using similar methodologies. Further comparisons of the NSYR data with 2002 U.S. Census data on households and with nationally representative surveys of adolescents—such as Monitoring the Future, the National Household Education Survey, and the National Longitudinal Study of Adolescent Health—confirm that the NSYR provides a nationally representative sample of U.S. teenagers aged 13 to 17 years and their parents without identifiable sampling or nonresponse biases (for details, see Smith & Denton, 2005). The follow-up sample that is used in the data sets comes from this initial sample of 3,290 teens. To obtain more information regarding the technical details and documentation of the NSYR, please visit http://www.youthandreligion.org/.

## A Note on Versions

All the commands and examples for this book were produced using Stata 13 for Windows. The primary commands and options are similar for older versions, dating back until at least Stata 9. There were, however, a few changes between Stata 11 and Stata 12 and then a few more with Stata 13. Most of these changes do not affect the actual functionality but rather deal with convenience and appearance. In fact, most of the substantive differences that the new users would encounter fall under the topics covered in Chapter 1.

Given the relatively minor changes between versions, many readers may still be using Stata 12 or even Stata 11. To address this potential challenge, this book includes three versions of the introductory material (i.e., Getting to Know Stata). The vast majority of the material in all versions is extremely similar, but all were included to prevent any confusion over the small dissimilarities.

For users of Stata 13, please start with Chapter 1: Getting to Know Stata 13. For users of Stata 12, please start with Appendix A: Getting to Know Stata 12, and then rejoin the book at Chapter 2. For users of Stata 11 (or older), please start with Appendix B: Getting to Know Stata 11, and then rejoin the book at Chapter 2. From that point on, all of the commands and strategies are equivalent across versions (although the appearance of the screenshots may be slightly different).

The vast majority of the commands presented are similar for Stata for Mac as well. The appearance and wording of some icons as well as the pathways for the point-and-click menus may be slightly different for a Mac operating system.

## A Note on Notation

Certain text in this book will be presented in a slightly different font. Generally, anything that you enter into or that comes out of Stata will be denoted with the

`typewriter` (i.e., Courier New) font. This font will be used to indicate variable names in a particular data set, such as `gender` or `ids`. It will also be used to show the display from the Stata Results window (if the actual screen shot is not shown).

This font will be used to denote a command that is entered into the Command window to perform a given operation. Additionally, if these commands are presented by themselves within a sentence, they will be set apart by a dash pre and post (e.g., `-replace-`) so that they are not confused with a variable name.

The majority of this book discusses the syntax command interface (i.e., the Command window) aspect of Stata. But there will be times when the menu, point-and-click interface is described. Menus (e.g., File), clickable buttons (e.g., OK), or keys on the keyboard (e.g., Enter) will be denoted with the Arial font.

Finally, Stata is a case-sensitive program, meaning that all commands and variable names must be typed exactly as they are shown. For the purposes of this book, this sensitivity means that at times the capitalization may not follow typical grammatical conventions. For example, if a variable name starts a sentence and that variable name is lowercase, then that sentence will start with a lowercase letter.

## References

Frankfort-Nachmias, C., & Leon-Guerrero, A. (2010). *Social statistics for a diverse society* (6th ed.). Thousand Oaks, CA: Pine Forge Press.

Smith, C., & Denton, M. L. (2005). *Soul searching: The religious and spiritual lives of American teenagers*. New York, NY: Oxford University Press.

# Acknowledgments

The author and SAGE Publications gratefully acknowledge the contributions of the following reviewers:

Karen Y. Holmes, *Norfolk State University, Norfolk*

Sean Kelly, *University of Notre Dame*

David Peterson, *Iowa State University, Ames*

Raymond Sanchez Mayers, *Rutgers University, New Brunswick*

Alexender Alexeev, *Indiana University*

Ryan C. Black, *Michigan State University*

Justin T. Denney, *Rice University*

Lisa Dilks, *West Virginia University*

Baodong Liu, *The University of Utah*

Daniel D. Partin, *University of Kentucky*

Parina Patel, *Georgetown University*

Garry Rolison, *California State University, San Marcos*

Winston Tripp, *University of West Georgia*

# *Brief Contents*

# PART I

## *Foundations for Working With Stata*

# 1

# Getting to Know Stata 13

For many people, learning any new computer software can be an anxiety-producing task. When that computer program involves statistics, the stress level generally increases exponentially. If you have similar feelings as you begin your journey into becoming a Stata user, do not fear, you are not alone. This book is designed with this apprehension in mind. One of the primary goals of this book is to help alleviate, or at least minimize, this anxiety as we move toward becoming effective and proficient Stata users. Keep in mind that at one time you may have had similar feelings about using e-mail or the Internet, and just as many people now feel extremely comfortable using these programs, by the end of this book, you will have a similar grasp of and comfort with Stata.

Before diving into all the details of using Stata, it is important to have an understanding of its various components. This chapter will serve as an introduction to the basic building blocks of the Stata program. Each of these aspects will be covered in much more detail throughout the book, but this chapter provides an overview of the basic functionality of the Stata program. The second section of the chapter explains how data are opened, imported, and entered.

## What You See[1]

When you open Stata, by double clicking on the Stata icon, for the first time, you will see the screen as shown in Figure 1.1.

[1]If you are using Stata 12, please use Appendix A: Getting to Know Stata 12 instead of this first chapter. If you are using Stata 11 (or Stata 10), please use Appendix B: Getting to Know Stata 11 instead of this first chapter. All the same features are covered, but Stata 13 has a slightly different appearance from these previous versions, which may make matching up what you see in the text and on your screen a bit confusing. Starting from Chapter 2, the vast majority of operations and commands are similar across versions. And the text specifically notes any particular features that are different for previous versions.