Preface

The main intent behind this book is *empowerment*: I want to help you to benefit from using Stata in your own research. Your research is probably demanding enough as it is, but to many researchers, the technicalities of data management and analysis can cause major problems—sometimes overwhelming problems. Stata has the tools you need; the purpose of this book is to help you use them.

Stata is a versatile program aimed at data management, statistical analysis, and graphics for research. It is dynamic, too, with new and improved tools being added by Stata monthly, and with contributions from an enthusiastic user community daily. This rapid development pace may make the inexperienced user feel a bit lost in what may initially look like a huge jungle. I want to help you become familiar with the basics and to benefit from some of the more advanced analytic tools. I will not be able to demonstrate everything Stata can do, but I hope to help you get started—and more.

This book is an introduction, written for the newcomer who has little or no experience with Stata. But it will also be a valuable companion for more advanced users. Although I wrote the book to meet the newcomer's needs, I chose to build it systematically, e.g., by putting everything about calculations in one chapter, from the basics to the more complex stuff. This systematic structure makes it easy to locate the information you need. Some of the exercises are aimed at beginners.

The systematic approach also means that you should not try to understand or learn everything in the sequence it is presented, e.g., in chapter 4 on command syntax. But now that you know it is there, when you have a general question on Stata's grammar, you can look in that chapter to find the answer.

The book's primary audience is people working with health research. When selecting which data management and analysis tools to demonstrate, I chose the tools that in my experience are most often used in health research. But there is much more to it than what is shown in this book. Official Stata has hundreds of commands; I selected a few of them and point to some other commands that might be useful to you. In addition to the official Stata commands, there are a thousand user-generated contributions. I point to a few of them, too, and demonstrate how to find and use them.

Writing this book has been a joy (mostly). One of the best parts of the experience has been the enthusiastic discussions I have had with people at StataCorp. In particular, Alan Riley, Vince Wiggins, and Bennet Fauber have given a lot of useful input, and Terri Schroeder and Lisa Gilmore have skilfully prepared the manuscript for printing. The most important input, however, was from the students I taught and supervised.

If you believe you have discovered an error, or if you have a suggestion for improving the book, please send an email to ishr@soci.au.dk.

Svend Juul

Aarhus, Denmark
February 2006

Online supplements

This book has several online resources associated with it, which you can find at

http://www.stata-press.com/books/ishr.html

Resources on this web site include

- Datasets.
- Programs, such as those for easy handling of output (see chapter 17).
- A do-file for each graph shown. I sometimes show only the minimal command needed to display a graph; the corresponding do-file includes all options used to obtain the final graph.
- A link to supplementary materials.
- Errors and corrections (if any) will be shown in an *Errata* section. If you believe you have discovered an error, or if you have a suggestion for improving the book, please send an email to ishr@soci.au.dk. Do not use this address to obtain help; for help, see chapter 2.
- There may be other resources placed on the web site after this book goes to press, so visit it to see what else appears there.

Notation in this book

Stata commands and the corresponding output are generally shown in these typefaces:

```
. webuse lowbirth.dta
(Applied Logistic Regression, Hosmer & Lemeshow)
. keep pairid low smoke
. list in 1/4, sepby(pairid)

     +-------------------------+
     | pairid   low   smoke |
     |-------------------------|
  1. |      1     0       0 |
  2. |      1     1       1 |
     |-------------------------|
  3. |      2     0       0 |
  4. |      2     1       0 |
     +-------------------------+
```

The commands you can enter are shown in boldface and are preceded by a period that represents Stata's command prompt. Do not type the period.

When referring to menu items, I use a sans serif font, such as

Statistics ▷ Summaries, tables, & tests ▷ Nonparametric tests of hypotheses

I use slant to show keystrokes, such as *Ctrl-C* and *Enter*.