



# Preface

S is a language and environment for data analysis originally developed at Bell Laboratories (of AT&T and now Lucent Technologies). It became the statistician's calculator for the 1990s, allowing easy access to the computing power and graphical capabilities of modern workstations and personal computers. Various implementations have been available, currently S-PLUS, a commercial system from the Insightful Corporation<sup>1</sup> in Seattle, and R,<sup>2</sup> an Open Source system written by a team of volunteers. Both can be run on Windows and a range of UNIX / Linux operating systems: R also runs on Macintoshes.

This is the fourth edition of a book which first appeared in 1994, and the S environment has grown rapidly since. This book concentrates on using the current systems to do statistics; there is a companion volume (Venables and Ripley, 2000) which discusses programming in the S language in much greater depth. Some of the more specialized functionality of the S environment is covered in *on-line complements*, additional sections and chapters which are available on the World Wide Web. The datasets and S functions that we use are supplied with most S environments and are also available on-line.

This is not a text in statistical theory, but does cover modern statistical methodology. Each chapter summarizes the methods discussed, in order to set out the notation and the precise method implemented in S. (It will help if the reader has a basic knowledge of the topic of the chapter, but several chapters have been successfully used for specialized courses in statistical methods.) Our aim is rather to show how we analyse datasets using S. In doing so we aim to show both how S can be used and how the availability of a powerful and graphical system has altered the way we approach data analysis and allows penetrating analyses to be performed routinely. Once calculation became easy, the statistician's energies could be devoted to understanding his or her dataset.

The core S language is not very large, but it is quite different from most other statistics systems. We describe the language in some detail in the first three chapters, but these are probably best skimmed at first reading. Once the philosophy of the language is grasped, its consistency and logical design will be appreciated.

The chapters on applying S to statistical problems are largely self-contained, although Chapter 6 describes the language used for linear models that is used in several later chapters. We expect that most readers will want to pick and choose among the later chapters.

This book is intended both for would-be users of S as an introductory guide

---

<sup>1</sup><http://www.insightful.com>.

<sup>2</sup><http://www.r-project.org>.

and for class use. The level of course for which it is suitable differs from country to country, but would generally range from the upper years of an undergraduate course (especially the early chapters) to Masters' level. (For example, almost all the material is covered in the M.Sc. in Applied Statistics at Oxford.) On-line exercises (and selected answers) are provided, but these should not detract from the best exercise of all, using **S** to study datasets with which the reader is familiar. Our library provides many datasets, some of which are not used in the text but are there to provide source material for exercises. Nolan and Speed (2000) and Ramsey and Schafer (1997, 2002) are also good sources of exercise material.

The authors may be contacted by electronic mail at

`MASS@stats.ox.ac.uk`

and would appreciate being informed of errors and improvements to the contents of this book. Errata and updates are available from our World Wide Web pages (see page 461 for sites).

#### *Acknowledgements:*

This book would not be possible without the **S** environment which has been principally developed by John Chambers, with substantial input from Doug Bates, Rick Becker, Bill Cleveland, Trevor Hastie, Daryl Pregibon and Allan Wilks. The code for survival analysis is the work of Terry Therneau. The **S-PLUS** and **R** implementations are the work of much larger teams acknowledged in their manuals.

We are grateful to the many people who have read and commented on draft material and who have helped us test the software, as well as to those whose problems have contributed to our understanding and indirectly to examples and exercises. We cannot name them all, but in particular we would like to thank Doug Bates, Adrian Bowman, Bill Dunlap, Kurt Hornik, Stephen Kaluzny, José Pinheiro, Brett Presnell, Ruth Ripley, Charles Roosen, David Smith, Patty Solomon and Terry Therneau. We thank Insightful Inc. for early access to versions of **S-PLUS**.

Bill Venables  
Brian Ripley  
January 2002



<http://www.springer.com/978-0-387-95457-8>

Modern Applied Statistics with S  
Venables, W.N.; Ripley, B.D.  
2002, XII, 498 p., Hardcover  
ISBN: 978-0-387-95457-8