The Basics of S-Plus (Fourth Edition)
Andreas Krause and Melvin Olson
Statistics and Computing
(2005) Springer, 444 pp
Price $59.95, ISBN 0-387-26109-5

This book provides a basic introduction to S-Plus statistical computing based on version 7.0 of S-Plus. It is suitable for people with little computing knowledge and basic understanding of statistical data analysis. Though this book is not as comprehensive as S-Plus manuals and all topics are not covered in great detail, using some simple examples, it nicely demonstrates many important features of S-Plus that are useful for statistical data analysis. Another nice feature of this book is a set of exercises at the end of each chapter with worked out solutions. This is helpful for self-study. A first time reader may be tempted to skip these exercises, but will be amply rewarded by reading them. Many new ideas and functions have been introduced in these exercises which are not there in the original text.

The book starts with an introductory chapter that contains a brief history of S, S-Plus and other related information. The next chapter is devoted to S-Plus graphical user interface (GUI) for Windows and Unix. One does not have to know any S-Plus syntax to go through this chapter, it requires only a little familiarity with easy 'point and click' methods of typical Windows software. This is particularly helpful for people from other disciplines who are interested in statistical data analysis but do not have much knowledge of programming.

The rest of the book has the primary focus on S-Plus language and it describes many things that cannot be done just using GUI. Chapter 3 gives the "first lesson" in this context. It contains some general information about S-Plus and some basic commands for arithmetic and logical operations. "Second lesson" in Chapter 4 goes beyond basic arithmetic operations, and it deals with different types of data structures like matrices, arrays, data frames and lists. It also gives an introduction to simple user defined functions and missing values. However, from Example 4.1 in pages 97-98, it is not very obvious how one can add row/column names to a matrix when names for the other dimension is already there. An example of this kind

may be helpful for the user. Also, there is a typo in the line 7 in Exercise 4.4, where y should be replaced by yy.

The next two chapters deal with S-Plus graphics. Chapter 5 gives an overview of traditional S-Plus graphics and it explains some basic commands related to the creation of graphs and their layout. Graphic options are explored in more detail in Chapter 6. This chapter is devoted to trellis graphics, which can be viewed as a graphical way of examining high-dimensional structure in the data by means of conditional one, two or three dimensional graphs for different sets of conditions on other variables.

Chapter 7 presents descriptive and graphical methods for exploratory data analysis. Functions related to standard statistical distributions and hypothesis testing are also introduced here. Chapter 8 introduces statistical modeling which mainly covers regression, ANOVA, logistic regression and survival analysis. These two chapters of the book are devoted to statistical analysis but unlike Venables and Ripley (2002), it is not comprehensive. Many important statistical methods like clustering, factor analysis, principle component analysis, tree models and time series remain untouched.

Next few chapters of this book are devoted to S-Plus programming. Chapter 9 introduces some basic programming techniques like operations on list structures, writing S functions and debugging, while Chapter 10 deals with object oriented programming described in Chambers (1998). Chapter 11 contains a discussion on different ways of reading and writing data and also the commands for reading the source code from a file. Chapter 12 gives some tips and tricks for efficient programming. It also describes the method for incorporating C and Fortran programs in S-Plus. A small subsection is devoted to factors, and it describes how factors, ordered factors and characters are treated differently by different S-Plus functions.

Chapter 13 gives a brief description on how S-Plus works under Unix and Windows. Chapter 14 lists down some sources for information on S-Plus. Finally, the book ends with a brief discussion on some similarities and differences between S-Plus and R programming language. However, instead of having a small chapter at the end it would have been better for R users if there had been a little more discussion on the unique features of these two languages and their differences. However, given the main emphasis of the book is on S-Plus, the authors cannot be blamed for it. Though the authors mentioned that most of the S-Plus codes of this book also run in R, for R users it will be better to consult other text books on R like Dalgaard (2002), Maindonald and Buran (2003) and Crawley (2005), which are missing from the reference list in Chapter 16.

Overall presentation in this book is good and this book is ideal for a

quick basic introduction to S-plus. It is well suited for self-study and may be considered as a text book for S-Plus statistical computing. If the reader carefully goes through this book, he will acquire substantial knowledge in S-Plus data analysis.

References

Chambers J.M. (1998). *Programming with Data : A Guide to the S-Language*, Springer-Verlag, New York.

Crawley, M.J. (2005). *Statistics : An Introduction using R*, Wiley, New York.

Dalgaard, P. (2002). *Introductory Statistics with R*, Springer-Verlag, New York.

Maindonald, J. and Buran, J. (2003). *Data Analysis and Graphics using R* Cambridge University Press, Cambridge.

Venables, W.N. and Ripley, B.D. (2002). *Modern Applied Statistics with S-Plus.* Springer-Verlag, New York.

Anil K. Ghosh

Centre for Mathematics and its Application

Australian National University, Canberra

ACT 0200, Australia

E-mail: anilkghosh@rediffmail.com