Conclusion

The **backtest** package provides a simple collection of tools for performing portfolio-based tests of financial conjectures. A much more complex package, **portfolioSim**, provides facilities for historical portfolio performance analysis using more realistic assumptions. Built on the framework of the **portfolio**[2] package, **portfolioSim** tackles the issues of risk exposures and liquidity constraints, as well as arbitrary portfolio construction and trading rules. Above all, the flexibility of R itself allows users to extend and modify these packages to suit their own needs. Before reaching that level of complexity, however, **backtest** provides a good starting point for testing a new conjecture.

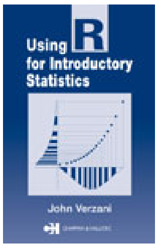
Bibliography

J. Enos and D. Kane. Analysing equity portfolios in R. *R News*, 6(2):13–19, MAY 2006. URL `http://CRAN.R-project.org/doc/Rnews`. 41

Kyle Campbell, Jeff Enos, Daniel Gerlanc and David Kane
Kane Capital Management
Cambridge, Massachusetts, USA
`Kyle.W.Campbell@williams.edu`, `jeff@kanecap.com`, `dgerlanc@gmail.com` *and* `david@kanecap.com`

Review of John Verzani's Book Using R for Introductory Statistics

Andy Liaw



To the best of my knowledge, this book is the first of its kind: a *standalone* introductory statistics textbook that integrates R throughout. The advantages should be obvious: Students would not need to purchase a supplement that covers the software, in addition to the main textbook (although the author states in the Preface that the book should also be useful as an accompaniment for a standard introductory text). That the software is freely available is a big bonus. Moreover, the book covers basic descriptive statistics before any probability models are mentioned. For students that are less mathematically inclined, this should make materials easier to absorb. (The author states in the Preface that the book aims at classes that are based on pre-calculus skills.)

The book contains 12 chapters. The first four chapters of the book cover descriptive statistics, both numerical and graphical, from general introduction (Chapter 1), through univariate and bivariate data (Chapters 2 and 3) to multivariate data (Chapter 4). Each chapter covers both categorical and numerical data. The author chose to treat two independent samples as bivariate data and several independent samples as multivariate data, which I think is a bit unusual. Chapter 5 covers probability models. Chapter 6 covers simulations, setting up for the topics on inference in the chapters that follow. Chapters 7 and 8 cover confidence intervals and significance tests, respectively. Chapter 9 discusses the χ^2 tests for the multinomial distribution, the test for independence, and goodness-of-fit tests such as Kolmogorov-Smirnov and Shapiro-Wilk. Chapter 10 covers both simple and multiple linear regression. Chapter 11 covers one- and two-way ANOVA as well as ANCOVA. Chapter 12 covers logistic and nonlinear regression. There are also five appendices that cover various aspects of R (installation, GUI, teaching with R, graphics, programming). Throughout the book, examples of R usage are interspersed among the main text, and some sections devoted to R topics are introduced as the need arises (e.g., in Chapter 6, Simulations, Section 6.2 covers `for()` loops). Data used as examples were drawn from a wide variety of areas. Exercises are given at the end of sections (rather than chapters). The book also has an accompanying add-on package, `UsingR` (available on CRAN), which contains data sets and some functions used in the book. The book also has a web site that contains answers to selected problems, the `UsingR` package for various platforms (including one for S-PLUS), as well as errata.

Several ideas presented in the book deserve accolades (e.g., covering EDA before introducing probability models, coverage of robust/resistant methods, thorough integration of R into the materials). However, there are also drawbacks. The most glaring one is the fact that many rather technical terms are used before they are introduced or explained, and some are not sufficiently elaborated. (E.g., "density" is first used to show how a kernel density estimate can be added to a histogram, but no explanation was given for what a density is or what it means.) In my teaching experience, one of the most difficult (but absolutely essential) concepts for students to grasp is the

[2] See Enos and Kane (2006) for an introduction to the **portfolio** package.

idea of the sampling distribution of a statistic, yet this did not receive nearly the attention I believe it deserves. The discussion of scatterplot smoothing (Section 3.4.7, "Trend lines") gave a minimal description of what smoothers are available in base R, such as smoothing splines, loess, and Friedman's supersmoother. I would be surprised if students in intro stat courses are not completely bewildered by such a minimal description. This will make it harder for some students to follow the text along on their own.

Some people might be interested in learning how to do basic data analysis using R. As these people are not among the intended audience, this book may not serve them as nicely as others, because the R-specific topics are scattered throughout the book in bits and pieces, each making its entrance as the statistical topic being covered requires it.

Now, some real nit-picking on more esoteRic things: The author seems to use "library" and "package" interchangeably, which could make some R users cringe. Also, on page 8, the virtue of using `<-` is touted, but the author still decides to use `=` for the book, without explanation. I also found the mention of the (evil?) `<<-` wholely unnecessary: The author said that it may be useful in programming R, yet personally I have not used it in the years I have been programming R. At the level the book is intended, I believe the students would be better served by its exclusion.

In summary, I like the structure of the book very much. However, the various problems mentioned above keep me from giving it a whole-hearted recommendation as a standalone text. It may serve well as a supplementary text for a more standard introductory Statistics textbook (*a la* Peter Dalgaard's "Introductory Statistics with R").

Andy Liaw
Merck Research Laboratories
`andy_liaw@merck.com`

DSC 2007

by Hadley Wickham

The fifth "Directions in Statistical Computing" conference was held in sunny Auckland, New Zealand, 8–10 February 2007. Despite the distant location, over 80 attendees from around the world made it to the conference, to enjoy the interesting talks, fabulous weather and the great food and drink of Auckland, including a fantastic ocean-side conference dinner.

The five key note speakers presented five quite different looks at the future of statistical computing:

- Ross Ihaka (re-)explored the use of LISP as a computational backend to R. He said, repeatedly, that this was just for fun, and will not be the next R (Q?)
- L. Fraser Jackson demonstrated the use of J (an open source grand child of APL) for statistics. A very concise and powerful language.
- Patrick Wessa gave an interesting talk about costs of the current scientific publishing framework and how we can do better.
- Olivia Lau discussed **zelig**, a modeling framework for R, which provides a common structure for modeling functions, making it easier for users and developers.
- Duncan Temple Lang talked how about a more composable object orientated R core (the code, not the people) would make it easier to experiment and explore new ideas, and continue to keep R relevant for the future.

A complete programme and abstracts are available on the conference website, `http://www.stat.auckland.ac.nz/dsc-2007/`. Conference papers will be published in a special issue of *Computational Statistics*.

A big round of thanks goes to the organizing committee, Paul Murrell, Ross Ihaka, David Scott and Thomas Yee, for such a great conference, as well as to Sharon Walker and Suryashobha Herle for feeding the hungry attendees.

Hadley Wickham
Department of Statistics
Iowa State University
Ames, Iowa, U.S.A.
`h.wickham@gmail.org`