Preface

Abstract MS Excel, the most widely available spreadsheet on MS Windows machines, is often used for data collection, manipulation, and storage. Elementary and medium-complexity mathematical and statistical functions are included with Excel. More advanced and highly reliable statistical analysis in Excel requires an add-in package. R is one of the best statistics programs available. It is an extensible system of software facilities for data manipulation, statistical analysis, and graphical display. With RExcel, the entire R environment (including more than a thousand contributed packages) can be treated as an extension of Excel.

This book is a supplementary text to any introductory course in statistics. The book supports the instructor by giving students step-by-step screenshots showing access to state-of-the-art statistical computations in R directly from the menu bar in Excel.

The book can also be used as a computational introduction by data analysts who already have basic statistical skills.

R is a program for statistical analysis and graphical display of data.

R is one of the best programs for statistical analysis and graphical display of data. It is maintained and distributed by an international team of statisticians and computer scientists working in universities and industry. R is one of the major tools used in statistical research and in applications of statistics in science, social science, economics, and business. R used in both academia and industry.

Among other things, R has

- data handling and storage facilities.
- a suite of operators for calculations on arrays, in particular matrices.
- a large, coherent, integrated collection of intermediate tools for data analysis.

- graphical facilities for data analysis and display either directly at the computer or on hardcopy.
- a well-developed, powerful, and effective programming language (called S) that includes conditionals, loops, user-defined functions (including recursive functions), functions for creating complex data structures, and input and output facilities. (Indeed, most of the system-supplied functions are themselves written in the S language.)
- A large selection of demonstration datasets used in the illustration of many statistical methods.

Excel is the most widely used spreadsheet program.

Microsoft Excel® [Microsoft, 2008a] is the most widely available spreadsheet. Entering data, cleaning data, and simple data processing (including simple statistics) are very easily done on spreadsheets. As a consequence, much statistical data is available as, or even created in, Excel worksheets.

Spreadsheets have a different paradigm for representing mathematical formulas than statistical (and mathematical) programming languages. The spreadsheet paradigm is much more visual and action-oriented than the functional or procedural paradigm of statistical programming languages. This problem of different paradigms can be overcome. In this book, we illustrate some of the ways the two paradigms can be made to work with each other. [Neuwirth and Arganbright, 2004] discuss in detail how to represent the development and structure of spreadsheets in printed form and how spreadsheets can be used to do serious mathematical work.

Excel is easy to use, but statisticians have found it has some deficiencies in the area of numerical precision. Statistical software is usually perceived as difficult to learn. This can be a major obstacle for potential users of advanced statistical methods. As this book shows, using R within Excel allows access to both the easy-to-use tools for data entry and manipulation available in Excel and the power and precision of the advanced statistical methods available via R.

RExcel is an interface program that uses R as an add-in to Excel.

RExcel is an add-in to Excel on MS Windows [Microsoft, 2008b] machines that allows the use of R as a "helper application" for Excel. Data can be transferred between Excel and R (in both directions), and Excel can call R functions to perform calculations and then transfer the results to Excel.

RExcel offers the following features:

- allows the use of R functions in Excel cell formulas, effectively controlling R calculations from Excel's automatic recalculation mechanism.
- connects R dataframes and Excel data lists.

- handles missing data.
- allows the creation of a standalone RExcel application that hides R almost completely from the user and uses Excel as the main interface to R. (Instructions are given in RExcel's help file.)
- if R Commander [Fox et al., 2007] is available, RExcel optionally places the Rcmdr menus on an Excel menu bar. Any menu item integrated into the Rcmdr menu using an Rcmdr plugin will also be available on the RExcel Rcmdr menu.
- works with the statconn (D)COM server (previously called R(D)COM server) server, turning R into an (invisible) background server for Excel.
- works with the rcom package, turning R into a (visible) foreground server for Excel. Using this configuration, the user can access the same instance of R either from Excel or from the command line in an R GUI Console window.
- supports R processes running under the control of RServerManager. R server is attached to Excel from a server pool. Different instances of Excel (running on different machines) may access the same R process with the same data. We do not use this capability in this book.

The RExcel interface is described in [Baier and Neuwirth, 2007]. RExcel is built on the the rcom and statconn (D)COM (previously called R(D)COM) packages, which we use for communication between R and the Microsoft Office software [Baier, 2007]. Basic information on the installation of R, RExcel, and Rcmdr is in Appendix A. Full information on RExcel is available at http://rcom.univie.ac.at/.

Rcmdr is an R package that provides GUI menu access to R.

Rcmdr (R Commander) is a platform-independent menu interface to R. The menu items implemented by Rcmdr are primarily designed for introductory courses. They can be extended by the Rcmdr plugin facility to provide a clickable graphical user interface (GUI) to any statistical procedure coded in R.

Audience

There are two audiences for this book:

1. students learning statistics.
2. people analyzing data.

Students

Introductory courses in statistics, and introductory statistics components of courses in all other subjects, require access to a software system for the collection and analysis of data.

This book is a supplementary text for an introductory course in statistics. We include examples of all the standard data analysis techniques that are introduced in such courses. We also include some of the elementary probability examples from those courses.

Many examples are structured parallel to similar presentations for other software that appears in such texts. The outline for such examples is as follows:

1. Read data into Excel from a textbook CD.
2. Put data into R from the RExcel menu.
3. Construct standard analysis tables and graphs from the Rcmdr menu installed in the Excel menu bar.
4. Cut and paste the tables and graphs into a document describing the results of the analysis.

We have two worksheets, using Excel and R only, that are used to illustrate the fundamental concepts of hypothesis testing, the construction and interpretation of confidence intervals, and the ideas behind least-squares fitting. We have several other worksheets that are used to illustrate data transfer between R and Excel and to illustrate additional statistical techniques.

Data Analysts

See the Students section above for an introduction to the use of RExcel and Rcmdr to access many of the analysis and graphical capabilities of R. It is possible to write additional menu items to access specialized functions written in R directly from the Rcmdr menu installed in the Excel menu bar. See the RExcel and Rcmdr documentation for details.

Updates and Additional Information

RExcel has a Wiki at http://rcom.univie.ac.at/.

Update material for this book will be available from the book's website at Springer http://www.springer.com/978-1-4419-0051-7.

A video on RExcel, including both the material in this book and additional material, is available at http://rcom.univie.ac.at/RExcelDemo/.

Acknowledgments

First and foremost, we have to thank Thomas Baier, without whose work [the rcom package and the statconn (D)COM server (previously called R(D)COM server)] RExcel and the book built on it would not have been possible. It should be noted that his design, now more than 10 years old, has not needed any change—a very uncommon event in the software world.

Christian Ritter has been the premier user of RExcel, and he has contributed many ideas to the design of the system.

We wish to acknowledge our students at Temple University and the University of Vienna who have used preliminary versions of this book and the RExcel software in class and made many helpful suggestions that have been incorporated into this version. We wish to thank Burt Holland at Temple University for teaching with an earlier version of the book and software. Our early experiences using RExcel in teaching are described in [Baier et al., 2006].

We wish to acknowledge John Fox of McMaster University both for the Rcmdr menu system [Fox et al., 2007] and especially for his willingness to incorporate changes into his system that were needed to make Rcmdr and RExcel cooperate. Rcmdr was designed as a platform-independent menu system. We have moved the Rcmdr menu to the Excel toolbar as part of our integration of Excel and R.

We wish to thank R Core for the R program [R Development Core Team, 2008].

Philadelphia and Vienna   *Richard M. Heiberger*
July 2009   *Erich Neuwirth*

R Through Excel
A Spreadsheet Interface for Statistics, Data Analysis,
and Graphics
Heiberger, R.M.; Neuwirth, E.
2009, XXIV, 344 p. 300 illus. in color., Softcover
ISBN: 978-1-4419-0051-7