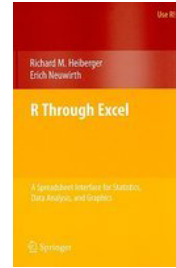


Book Review:

## **R through Excel: A spreadsheet interface for statistics, data analysis, and graphics**

by Richard M. Heiberger & Erich Neuwirth  
New York: Springer Science+Business Media (2009)  
ISBN 978-1-4419-0051-7



*R* (<http://www.r-project.org>) is a robust, speedy, and convenient cost-free environment for statistical computing and graphics creation. It includes a programming language and is supported by an active and generous community that has produced any number of convenient pre-written code batches for any number of widely used statistical functions (as well as a great many that *should* be widely used). Unfortunately, *R* comes from an academic environment unfamiliar to many working in language testing. It often necessitates learning a daunting number of new terms for familiar things, new ways of arranging, manipulating, and storing data, and even new ways of referring to results. The book *R Through Excel* and the software packages *RExcel* and *(D)COM* (Baier & Neuwirth, 2007) offer a good entry for those just beginning with *R* through the familiar Microsoft Excel. This text will help those using Microsoft Excel on computers in MS Windows environments become more familiar with two programs designed to work with *R*. However, all of the functions and many of the topics discussed in the book can be applied for independent *R* installations on Linux, Unix, or Apple computers.

### **Preliminaries and Appendices**

*R Through Excel* begins with front matter that readers would do well to skim. Additional to a justification for the book and information on its intended audience are pleasingly stern admonitions regarding rounding and use of graphics. Information about how to download and install *R* and various related programs are included. There are also tips for using *R* in non-Windows operating systems or the OpenOffice system, and commonly encountered error messages as well as ways to respond to them in the appendices.

Most readers will find themselves turning to Appendix 1 first, as it contains instructions for installing *RExcel*, *RCommander* (an interface for *R* that is slightly less intimidating than its native command window), and *(D)COM*.

### **Chapters 1 - 3**

How to start the *RExcel* and *RCommander* programs are covered for various usage situations (*e.g.*, while Excel alone is running or while *R* alone is running). Again, this information is useful for those new to the programs. The written information is backed with clear, color illustrations covering both Excel 2003 and 2007.

Chapter 2 then offers a simple tour of the *RExcel* menus, interspersed with useful pieces of information, including ways to get data into Excel from *R* and how to get help for *R* functions. This is a good place for *RExcel* beginners to start reading closely.

The most useful part of the text begins at Chapter 3, when the authors detail how to get data from *R* into Excel and vice versa. As an often frustrated user of *R* who stores data in Excel, the capabilities of *RExcel* are perhaps the most welcome aspect of the program. Data must be formatted for *R* in ways far different from what most users are used to. Novice or occasional *R* users might find understanding data formats frustrating and time-consuming enough to give up on the program. *RExcel* allows relatively painless data transfer from Excel to *R* and back. The word “relatively” is warranted: I encountered difficulties when using one of the two transfer methods. These difficulties are not addressed in this otherwise thorough book, leading me to believe they’re affected by the software environments on the two roughly similar computers I use. Fortunately, the authors (and *RExcel*) mention more than one data transfer method. By following the ample illustrations readers should be able to work through most quirks.

### **Chapters 4 and 5**

Judging from the titles (“Normal and *t* distributions” and “Normal and *t* workbooks”), these chapters might seem worth skipping, but I do not recommend this. Not only is the authors’ explanation of the normal and *t* distributions simple and relevant to much of what language testing research, but the chapters also cover calculating *p* values and statistical power (addressing *alpha* and *beta* concerns). Those educated in language testing or familiar with *SPSS* are well advised to peruse the chapter carefully. This is because much of the

terminology in applied linguistics differs from what is used in the text. For example, what most people know as a “scatterplot” is described in *R* as an “X-Y conditioning plot.” Without recourse to dictionaries, the only way to familiarize oneself with the terms used in the examples in *R Through Excel* is to read the examples thoroughly and compare the terms used with those one is familiar with.

## Chapters 6 - 9

How to do *t*-tests, one-way ANOVA, and linear regression are covered in these chapters. Those comfortable with calculating the statistics for these may well skim this section, especially if they plan to continue using other programs for such analyses. However, much valuable information on matters not limited to *t*-tests, ANOVA, or regression is also covered here. For example, the *t*-test chapter contains a useful discussion of confidence intervals, what they mean, and how to calculate them. Topics such as multiple comparisons, regression, confidence and prediction bands are well elucidated.

Chapter 9 is given over to a discussion of least squares and, importantly, residuals and leverage. It is not entirely clear why this chapter is placed here, between a chapter on (simple) linear regression and one on multiple regression, but the discussion, illustrations, and accompanying demonstration spreadsheet are useful as a review or introduction.

## Chapters 10 – 13

Here the authors cover multiple (dual *x*), polynomial, and multiple (3 or more *x*) regression, respectively. Readers seldom working with such models may want to skim these chapters, but they should not be skipped, as the description of model fit statistics is useful.

Like the rest of the chapters in the latter half of the book, the final chapter is both a computing guide and discussion of a family of statistical procedures. This chapter covers contingency tables and chi-square tests. Though the chapter is short, it contains clear examples and a lucid explanation of chi-square residuals.

## A Few Shortcomings

This text is relatively short, comprising about 350 pages of text loaded with illustrations, and printed on dull coated offset paper and hence is relatively expensive: ¥7,817 is the standard price at Amazon.co.jp. Those unhampered by a tight budget might opt instead for *The R Book* (Crawley, 2007). Though more costly (¥8,602 at the time of writing), this text offers a more comprehensive introduction to *R* along with fine explanations of a very wide range of statistics. With patience, the information conveyed in *R Through Excel* can be puzzled out by novices simply using the *R* manuals and the *RExcel*, *RCommander*, and *(D)COM* programs (all free), but my experience is that the process would be lengthy and often frustrating. If you can afford only one, a text such as *The R Book* is a better option. Buying both books is the best option for researchers with enough money.

It is well that *R Through Excel* is so short, since its limited index makes it hard to navigate: simply paging through the contents is more likely to take readers where they want to go. The relative brevity is, again, a relief given the organization of the information in the book: it would have been nice, for example, to find all of the information about producing graphics either in the same place or summarized with index entries.

Regarding content, the one issue I took issue with was the authors’ emphasis on formatting numbers with decimals aligned in an Excel format. The *RExcel* program includes a context menu item for formatting numbers so that the decimals align, but I do not find numbers with unaligned decimal points particularly difficult to read or understand.

Another welcome change would have been more information on calling *R* functions in Excel cells: *RExcel* with *(D)COM* allows this, but the authors’ mention it only in passing. Granted, the authors concentrated on the *RExcel* program they wrote, but a focus suited to more novice users would have been welcome.

## The Final Word

I recommend this book for anyone working with even the simplest statistics or graphs, simply because the *RExcel* program allow such an easy introduction to *R* and the book explains *RExcel* well.

The style is straightforward and, indeed, at times amusingly stern, as when they lay down the law on correct and incorrect graphics and tabular formatting.

The *RExcel* and *(D)COM* programs cannot be recommended highly enough. In 2005, J. D. Brown published his admirable *Testing in Language Programs*. Pages 190 to 192 show the reader how to calculate Cronbach’s alpha in Excel and pages 193 to 195 specified how to calculate the K-R20 measure of internal consistency reliability. This can be done in Excel using *R* and *RExcel* with a single formula: `RApply` (“alpha”,

*cells*), where *cells* is the cell range of the test data. Rasch and other probabilistic approaches to test analysis as well can, with *R*, be run more or less directly in Excel. *R* graphics far outstrip the primitive and, at times, misleading, charts produced by Excel natively.

*R* is an intimidating but powerful program that assumes an intimate knowledge of data formats and terminology not familiar to many language testers. *R Through Excel* is a highly recommended first step into that program.

-Reviewed by Hiram Bucheem

### **References**

Brown, J. D. (2005). *Testing in language programs*. New York: McGraw Hill.

Crawley, M. J. (2007). *The R book*. Chichester, West Sussex, England: Wiley.

Baier, T., & Erich Neuwirth, E. (2007). Excel :: COM :: R. *Computational Statistics*, 22 (1), 91-108.