Book Review

Reviewed by Zhongxian Wang* and James Yao

Email: wangz@mail.montclair.edu
Email: yaoj@mail.montclair.edu
*Corresponding author

Managing Data Using Excel: Organizing, Summarizing and Visualizing Scientific Data
by: Mark Gardener
Published 2015
by Pelagic Publishing
P.O. Box 725, Exeter, EX1 9QU, UK, 326pp
ISBN: 978-1-78427-007-0

Hans-Jürgen (2015) outlines the main features of present-day statistical software by reviewing its development from the (late) 1950s to the present time. Out of so many available on-shelf software in data handling, Excel is the most popular, painless, and powerful package. Many of us have heard about this spreadsheet application, but are still not very accustomed to it as we are to Microsoft Word. However, we can all take full advantage of Excel by reading *Managing Data Using Excel: Organizing, Summarizing and Visualizing Scientific Data* by Mark Gardener.

This book provides valuable and helpful advice to users of any age, stage, and level, including experts, to analyse and calculate data seamlessly. The author explains in vivid details, comprehensible to anyone, all the how-to's of Excel. The first part of this book deals with the preparation, arrangement and management of your dataset, and the second part focuses on the use of your dataset. Moreover, it discusses options to dig and share your data and analysis.

Part 1 Arranging and managing your data

In order to manage your data more easily and efficiently, you should prepare your data and arrange them in an appropriate way. A significant part of this data management contains editing, rearranging, and check for errors.

Part 1 consists of three chapters. Chapter one demonstrates *scientific recorking format*, a sensible and systematic approach to data layout. Setting out your data in a logical and useful manner is an extremely important starting point. Users often underestimate the importance of this aspect. The scientific process can be split into four parts: planning, recording, analysing and reporting. The laid out data should show all the information, be flexible for extension and modification, and be easy for checking.

Chapter two shows the building of your dataset and embraces a range of data management issues, such as adding, editing, and rearranging your data. In *scientific*

recorking format, each column represents a separate variable and each row is a record or observation. Manufactured variables can take various forms: index variables, date variables, miscellaneous variables (*binary*, *categorical*, *label*, *replacement*, and *abbreviation*). Also, you can add notes to your data. You can sort, filter, and rearrange your data.

Chapter three, the last chapter of Part 1, presents management issues relating to error checking, an often overlooked task. Error checking should not be postponed until you have finished collecting and entering your data. Spotting errors at an early stage can save or rescue you. You can constantly check potential errors by validating data as it is entered because spelling errors and general typos in text entries are almost inevitable. Furthermore, errors with numerical data are more harmful and can distort your research findings.

Part 2 Using your dataset – summarising, visualising and sharing your data

Part 2 (chapters 4–9) tracks ways of using your data. Chapter four overviews the methods and tools available for users. Data can fall into one of the three main camps: correlation and regression; association; and differences. Summary statistics come in four main sorts: averages (middle values), variability, replication, and distribution (data frequency and the shape of the distribution). The *Analysis ToolPak* is a Microsoft add-in for Excel.

Chapter five emphasises exploring regression data. Looking up for relationships between numeric variables depends on regression analysis. According to the author, correlation is a simpler form of regression and is concerned with both the strength and direction of the link between two numeric variables. Also, you may look at the differences between grouping categorical variables. There are two main ways to explore your data: numerically with correlation matrices; and visually with scatter plots and other kinds of chart.

In line with the author, each kind of dataset lends itself to a different way of being explored. Chapter six starts to explore time-related data. A pivot table is essential for most time-related data that allows users to rearrange their data in a sensible manner. Users need to split the time into smaller chunks if many observations have different time periods. *Sparklines* and *Line plots* are two main types of plot users can produce. Users can either directly or indirectly make a pilot chart.

Chapter seven focuses on exploring association data. Input data would take a particular form: counts of items that fall into particular categories. A contingency table could be the end result. The values in the contingency table display the frequency, the number of items in each of the combinations of categories. You cannot use averages and the usual ways of summarising numerical samples because association data are not replicated data. *Stacked charts* have two categories: *regular* and *percentage*.

Three basic measures are needed for summarising any sample: centrality, dispersion, and replication. For overcoming the problem of hiding the unwanted cells, the author suggests to use the advanced filter in chapter eight, which allows you to send your filter results to a new worksheet. The pivot table is a crucial element in exploring difference data because it allows users to group and re-group their data efficiently, as well as to visualise data via a pivot chart.

Chapter nine talks about sharing your data. Hypotheses are tested by collecting data and exploring the results. Scientific knowledge on a certain topic is gained. A scientific

research project will not be completed until the results to targeted recipients and/or general public is presented and shared. In some cases, users probably need to share their data with another program, like R (Gardener, 2012), for a deeper analysis. In addition to a main worksheet with the actual data, recipients may need other worksheets, such as notes and methodology, pivot tables, charts, lookup tables, etc.

Overall, *Managing Data Using Excel* is an indispensable book for all students, researcher, and managers who use data and are seeking to manage data more effectively. It is aimed at problem solvers and decision makers at all levels, but it is especially useful for university-level research, from undergraduates to postdoctoral researchers. In addition to the crystal-clear text, highlighted important points, helpful self-assessment exercises, and summary tables can be found at the end of each chapter. Furthermore, Have a Go exercises, Tips and Notes give readers practical experience. Many supplementary materials can also be downloaded from the companion website.

Excel is an extremely powerful tool because almost every scientific discipline depends on statistical analysis of experimental data. Statistics plays a predominant role in almost all natural and social science studies. A well-founded understanding of basic statistics is critical for everyone in science to produce statistically sound research. However, analysis of statistics could become an entangling snare since there are so many pitfalls in statistical analysis. As Reinhart (2015) pointed out, "You would be surprised how many scientists are doing it wrong." The first step toward statistics done right is *Statistics Done Wrong*.

This book can help users, from high school students to university-level students to scientists at all levels, maximise the usefulness of their data. By the way, this book belongs to research skills series that enable readers to develop skills and confidence in academic research; pursue in-depth investigation and analysis; produce consistent and effective reporting analysis; and build confidence in working with large datasets (http://www.pelagicpublishing.com).

References

Gardener, M. (2012) *Statistics for Ecologists Using R and Excel: Data Collection, Exploration, Analysis and Presentation (Data in the Wild)*, Pelagic Publishing, Exeter, UK.

Hans-Jürgen, A. (2015) 'Statistical analysis packages', in Wright, J.D. (Ed.): *International Encyclopedia of the Social & Behavioral Sciences*, 2nd ed., pp.376–380, Elsevier Ltd., Oxford.

Reinhart, A. (2015) *Statistics Done Wrong: The Woefully Complete Guide*, 1st ed., No Starch Press, San Francisco, CA 94103, USA.