# A Critical Review of How Well Reference Manuals and Textbooks Inform The Reader about Concerns with Using Microsoft Excel 2000 for Statistical Analyses

John D. McKenzie, Jr.
*Babson College, Math/Science Division*
*231 Forest Street*
*Babson Park, USA*
*mckenzie@babson.edu*

## 1. Introduction

Microsoft Excel 2000 is used for statistical analyses by a large number of academics and practitioners. Most of these professionals and their students are unaware that are numerous problems with using this popular spreadsheet package for such analyses. It is now known that the package contains poor algorithms for linear regression (McCullough and Wilson (1999)). It is also known that some of its formulas are wrong, e.g., its formulas for the first and third quartiles. There are also serious concerns with the user-friendliness of the package. It requires the predictors in a multiple linear regression to be contiguous. It does not handle missing data well. In addition, the package does a poor job at providing help to the user. In 2001 the author and two of his colleagues (William Rybolt and Elaine Allen) made a presentation on 21 different concerns with using Excel for statistics. They also noted that most of these concerns exist in all releases of Excel that contain its Data Analysis ToolPak.

## 2. Textbooks

One would hope that authors would acknowledge these concerns if they decided to incorporate Excel into their textbooks. But from an examination on 13 textbooks for a presentation made by Allen, McKenzie, and Rybolt in 2000 this was not the case. A few authors cautioned the reader in the preface about the use of Excel or mentioned a specific weakness in the body of the text, but most did nothing. In this paper the author will examine the latest editions of these and other textbooks that incorporate Excel to see how many of the 21 concerns are mentioned in each textbook.

## 3. Reference Manuals

Another source of information about the use of Excel for statistical analyses are the reference manuals devoted to this subject. In this paper the author will critique how well five of the most

popular such manuals inform the reader of the concerns mentioned above. These manuals are Berk and Carey (2000), Dretzke and Heilman (1998), Eldredge (1999), Middleton (2000), and Neufeld (2001). In addition he will produce a comparative review of these five manuals.

**4. Conclusion**

At the end of the paper the author will report on a 2001 Joint Statistical Meetings panel discussion on the use of Excel 2000 for statistical analyses that he jointly organized with Jon Cryer. Included in this summary will be the remarks from an author deliberating on how to incorporate Excel in a new textbook, from McCullough on the generation of Excel's random numbers, and from a Microsoft representative who is leading an effort to correct some of the algorithmic problems in Excel.

**REFERENCES**

Berk and Carey (2000). Data Analysis with Microsoft Excel, Updated for Microsoft Office 2000. Duxbury. Pacific Grove, California.

Dretzke B. J. and Heilman K. A. (1998). Statistics with Microsoft Excel. Prentice Hall. Upper Saddle River, New Jersey.

Eldredge D. L. (1999). An Excel Companion for Business Statistics. South-Western College Publishing. Cincinnati.

McCullough B. D. and Wilson B. (1999). On the accuracy of statistical procedures in Microsoft Excel 97. Computational Statistics & Data Analysis 31, 27-37.

Middleton M. R. (2000). Data Analysis Using Microsoft Excel, Updated for Office 97. Duxbury. Pacific Grove, California.

Neufeld J. L. (2001). Learning Business Statistics with Microsoft Excel 2000. Prentice Hall. Upper Saddle River, New Jersey.

**RESUME**

Cet article critique des manuels et les livres de classes pour leur traitement des analsyses statistiques pour Excel 2000. Il examine en détail cinq manuels: Berk and Carey (2000), Dretzke and Heilman (1998), Eldredge (1999), Middleton (2000), and Neufeld (2001).