

Contents lists available at [ScienceDirect](http://www.sciencedirect.com)

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Editorial

Special section on Microsoft Excel 2007

“[I]t is quite possible that more basic statistical calculations are done worldwide in EXCEL than in all statistical packages combined”. – [Wilkinson \(1994\)](#)

“Let’s not kid ourselves: the most widely used piece of software for statistics is Excel”. [Ripley \(2002\)](#)

At least since [Sawitzki’s \(1994a; 1994b\)](#) project, which was published in the pages of this journal, it has been known that serious errors exist in Excel’s statistical procedures. As has been recounted in these pages since then, Microsoft has not been very good at fixing the errors, often not bothering to fix known errors and sometimes introducing new errors while attempting to fix old ones. Consequently, when a new version of Excel is released, everyone wants to know, “Has Microsoft fixed the errors?” Papers and notes on various aspects of Excel’s statistical functionality might appear scattered in various outlets over time. For the release of Excel 2007, the special section has sought to collect the various papers and publish them together.

David Heiser, author of the website, “Microsoft Excel 2000, 2003 and 2007 Faults, Problems, Workarounds and Fixes” ([Heiser, 2006](#)) and [McCullough](#) have written the first paper ([McCullough and Heiser, 2008](#)), in which they assess some statistical procedures in Excel 2007, reprising [McCullough and Wilson \(2005\)](#). However, they also report errors in several functions that [McCullough and Wilson](#) never considered. For example, the “normal probability plot” is customarily used to compare regression residuals to a normal distribution; in Excel, it compares the dependent variable to a uniform distribution.

In the second paper, [Yalta \(2008\)](#), who has published on accuracy in other journals (e.g., [Yalta \(2007\)](#) and [Yalta and Yalta \(2007\)](#)), reviews the checkered history of Excel’s statistical distributions and examines Excel 2007. Microsoft has on more than one occasion “fixed” a problem by replacing one inaccurate algorithm with a different inaccurate algorithm, and often fails to make any fixes at all. He finds that Microsoft still has not fixed all the errors in the Excel statistical distributions.

In Excel 2003, Microsoft attempted to implement the Wichmann-Hill random number generator (WH-RNG). Instead of producing numbers between zero and unity, it would also produce negative numbers. Microsoft soon issued a patch; but how can users tell that the patch is correct? Did Microsoft just trap the negative numbers or did it correctly implement the WH-RNG? This problem is tackled in the third paper by [McCullough \(2008\)](#). He finds that Microsoft has again failed to implement correctly the Wichmann-Hill RNG.

Graphics packages for creating advertisements or pretty patterns have their own purposes. Graphics for displaying data fall under the discipline of statistical graphics, which has its own membership organization (ASA Section on Statistical Graphics), books, and journal (*Journal of Computational and Graphical Statistics*). One would hope that Microsoft would avail itself of this learned material, but Microsoft has resisted the temptation.

Stephen Few, a consultant who specializes in data visualization for business and who has written books on the subject recently wrote of a 2005 correspondence he conducted about Excel 2007 graphics ([Few, 2006](#)):

I learned who at Microsoft was responsible for the development of the new charting engine and quickly sent him an e-mail. In it, I introduced myself and offered to send him my suggestions for shaping Excel’s new charting functionality. With his polite assent, I submitted a list of recommendations. In response, I received a very courteous e-mail. Here’s an excerpt:

Thanks. This was quite interesting and useful. I think you’ll be pretty happy with some of the changes we are making for Office 12. I did read your book for inspiration (as well as Tufte, Cleveland, Wilkinson, Zelazny, etc).

So Microsoft is aware of this body of knowledge. In the fourth paper, [Su \(2008\)](#), who studied statistical graphics under Andrew Gelman at Columbia University, reports that Microsoft has chosen to ignore the precepts of statistical graphics. He examines the extent to which Excel’s graphical defaults facilitate or hinder the production of good graphics, and concludes

that the defaults lend themselves to chartjunk. It is possible to make good graphs by invoking the appropriate options, but to invoke these appropriate options the user would have to be well-versed in statistical graphics.

The fifth paper, on teaching statistics with Excel, is written by Nash (2008). Nash is a professor at the University of Ottawa, where he has taught statistics, forecasting and managing technological risk since 1980. His books, articles and papers cover many topics, but statistical computing is a common theme. Having built and used computers since the 1960s, he took an early interest in spreadsheets, and has published several articles on the uses of spreadsheets. He often uses spreadsheets in his courses. Currently he is working on the open-source “TellTable” project to provide an audit trail for spreadsheets, as well as another project for testing spreadsheets.

Over the years, this journal has published many articles describing the errors in Microsoft Excel, errors that raise concerns about Microsoft’s quality assurance procedures, at least insofar as Excel is concerned. Hence it is perhaps worth comparing quality assurance in Microsoft’s Excel to quality assurance in its game division. A recent issue of *Wired* magazine (Thompson, 2007) describes the way that “Bungie”, a Microsoft gaming subsidiary, tests one of its products, a game called Halo3:

Because it [Bungie] is owned by Microsoft, which launches dozens Xbox and PC games every year, Bungie has access to one of the most advanced game-testing facilities ever built. [Bungie has] now analyzed more than 3000 hours of Halo3 played by some 600 everyday gamers, tracking everything from favored weapons to how and where – down to the square foot – players most frequently get killed.

Bungie doesn’t just test its own games this way. It also buys copies of rival titles and studies those, too, to see how Halo3 matches up.

It is difficult not to think that if Microsoft tested business software the way it tested game software, then the statistical functions in Excel would be as accurate as those found in any other major software package. If that were the case, then none of the articles in this special section would have been written.

References

- Few, Stephen, (2006). Excel’s new charting engine preview of an opportunity missed, Business Intelligence Network, UK edition, 9 May 2006. <http://www.b-eye-network.co.uk/view-articles/2818> (accessed 30.08.07).
- Heiser, David A., 2006. Microsoft excel 2000 and 2003 faults, problems, workarounds and fixes. *Computational Statistics and Data Analysis* 51 (2), 1442–1443.
- McCullough, B.D., 2008. Microsoft excel’s ‘not the Wichmann–Hill’ random number generators. *Computational Statistics and Data Analysis* 52 (10), 4587–4593.
- McCullough, B.D., Heiser, David A., 2008. On the accuracy of statistical procedures in microsoft excel 2007. *Computational Statistics and Data Analysis* 52 (10), 4570–4578.
- McCullough, B.D., Wilson, Berry, 2005. On the accuracy of statistical procedures in microsoft excel 2003. *Computational Statistics and Data Analysis* 49 (4), 1244–1252.
- Nash, John, 2008. Teaching statistics with excel 2007 and other spreadsheets. *Computational Statistics and Data Analysis* 52 (10), 4602–4606.
- Ripley, Brian D., (2002). Statistical methods need software: A view of statistical computing, Opening lecture RSS 2002, Plymouth.
- Sawitzki, Gunther, 1994a. Testing numerical reliability of data analysis systems. *Computational Statistics and Data Analysis* 18 (2), 269–286.
- Sawitzki, Gunther, 1994b. Report on the reliability of data analysis systems. *Computational Statistics and Data Analysis* 18 (2), 289–301.
- Su, Yu-Sung, 2008. It’s easy to produce chartjunk using microsoft excel 2007 but hard to make good graphs. *Computational Statistics and Data Analysis* 52 (10), 4594–4601.
- Thompson, Clive, 2007. The science of play. *Wired* 15 (9), 140. et seq.
- Wilkinson, Leland, 1994. Practical guidelines for testing statistical software. In: Dirschedel, P., Ostermann, R. (Eds.), *Computational Statistics*. Berlin, Physica-Verlag.
- Yalta, A.Talha, 2007. The numerical reliability of GAUSS 8.0. *The American Statistician* 61 (3), 262–268.
- Yalta, A.Talha, Yalta, A.Yasemin, 2007. GRETL 1.6.0 and its numerical accuracy. *Journal of Applied Econometrics* 22 (4), 849–854.
- Yalta, A.Talha, 2008. The accuracy of statistical distributions in microsoft excel 2007. *Computational Statistics and Data Analysis* 52 (10), 4579–4586.

B.D. McCullough*
LeBow College of Business,
Department of Decision Sciences,
Drexel University,
Philadelphia, PA, 19104-2875,
United States
E-mail address: bdmccullough@drexel.edu.
 Available online 12 March 2008

* Tel.: +1 215 895 2134.