Issues With Using Microsoft Excel

For Statistical Analysis and Graphics

Bell CE. Excel Statistical Functions. *Proceedings of the Annual Meeting of the American Statistical Association*, August 5-9, 2001
[NOTE: Colin Bell was hired by Microsoft to "fix" Excel]
Top 10 list of problem areas:

1. Sums of squares
2. Multiple linear regression
3. Standard normal function (NORMSDIST)
4. Inverse functions for continuous distributions (CHIINV, FINV, NORMSINV, TINV)
5. Random number generator (RAND)
6. Numerical overflow for discrete probabilities (BINOMDIST, POISSON, HYPGEOMDIST)
7. Functions (PERCENTILE)

Helsel, D.R. (2002), Is Microsoft Excel an Adequate Statistics Package?
http://www.practicalstats.com/Pages/excelstats.html.

1. Commonly-used statistics and methods not available within Excel

2. Several Excel procedures are misleading.
3. Distributions are not computed with precision.
4. Routines for handling missing data were incorrect.
5. Regression routines are incorrect for multicollinear data.
6. Excel requires X variables to be in contiguous columns.
7. Ranks of tied data are computed incorrectly.
8. Many of Excel's charts violate standards of good graphics.

McCullough BD (1998). Asssessing the reliability of statistical software: Part I. *The American. Statistician* 52: 358-366.
McCullough BD and Wilson B (2002). On the accuracy of statistical procedures in Microsoft Excel 2000 and Excel XP. *Computational Statistics & Data Analysis* 40: 713 - 721.
Collection of intermediate-level tests to assess numerical reliability of a statistical software package in:

1. Estimation (both linear and nonlinear)
2. Random number generation
3. Statistical distributions (e.g., for calculating p-values)

- "Excel's performance in all three areas is found to be inadequate. Persons desiring to conduct statistical analyses of data are advised not to use Excel."
- "On the basis that Excel implements an unreliable algorithm for computation of the sample variance, its performance on this suite of tests can be judged inadequate. Sawitzki (1994b) noted that Excel 4.0 had the same difficulty calculating the sample variance, so Microsoft did not fix this error."
- All problems found in Excel 97 are still there in Excel 2000 and XP. "Microsoft attempted to fix errors in the standard normal random number generator and the inverse normal function, and in the former case actually made the problem worse."

Using Excel for Statistical Data Analysis, Eva Goldwater
Data Analysis Group, Academic Computing, U Mass,1997; updated 1999
Concluded that "Excel is a poor choice for statistical analysis beyond the simplest descriptive statistics, or for more than a very few columns."
Problems

1. Missing values handled inconsistently, incorrectly.

2. Data organization differs according to analysis, forcing reorganization of data for different analyses.
3. Analyses done on one column at a time
4. Output poorly organized, inadequately labeled, no record of how an analysis was accomplished.

Statistical analysis using Microsoft Excel. Jeffrey Simonoff, (2002) at http://www.stern.nyu.edu/~jsimonof/classes/1305/pdf/excelreg.pdf
Regression output in Analysis Toolpak of Microsoft Excel 2002

….output suppressed….

"Each of the nine numbers given above is incorrect! The slope estimate has the wrong sign, the estimated standard errors of the coefficients are zero (making it impossible to construct t–statistics), and the values of R^2, F and the regression sum of squares are negative! … Unless Excel does better at addressing these computational problems, it cannot be considered a serious candidate for use in statistical analysis".

Knut M. Wittkowski, PhD, DSc, The Rockefeller University Hospital, May 7, 2002
"Excel was designed for accounting and it should be used only for accounting, unless you have to deal with numbers above $10,000,000, in which case you should be able to afford software that is reliable."

What to use instead of Excel

- SAS
- S
- Splus
- R (www.r-project.org)
- JMP
- SPSS
- NCSS
- Systat
- Minitab
- StatPlus (Excel add-in routines), with textbook, "Data Analysis with Microsoft Excel".

- adequate, though still not in the areas of multiple regression and ANOVA for more than one factor
- Open Office or Star Office (www.sun.com)
- Any other statistical package!

REFERENCES

Wittkowski K.M. MS Excel for Data Management and Statistical Analysis? ***Association of General Clinical Research Center Statisticians Annual Meeting.*** **Atlanta, GA, August 4-5, 2001 (oral presentation).**

Helsel, D.R. (2002), Is it practical to use Excel for stats?, (http://www.practicalstats.com/Pages/excelstats.html).

(1) On the accuracy of statistical procedures in Microsoft Excel 2000 and Excel XP B.D. McCullough and B. Wilson, (2002), Computational Statistics & Data Analysis, 40, pp 713 – 721

(2) On the accuracy of statistical procedures in Microsoft Excel '97 B.D. McCullough and B. Wilson, (1999), Computational Statistics & Data Analysis, 31, pp 27-37 http://www.elsevier.com/gej-ng/10/15/38/37/25/27/article.pdf

(3) Problems with using Microsoft Excel for statistics [pdf Download] J.D. Cryer, (2001), presented at the Joint Statistical Meetings, American Statistical Association, 2001, Atlanta Georgia at http://www.stat.uiowa.edu/~jcryer/JSMTalk2001.pdf

(4) Use of Excel for statistical analysis. Neil Cox, (2000), AgResearch Ruakura at http://www.agresearch.cri.nz/Science/Statistics/exceluse1.htm

(5) Using Excel for statistical data analysis. Eva Goldwater, (1999), Univ. of Massachusetts Office of Information Technology http://www.umass.edu/acco/statistics/handout/excel.html

(6) Statistical analysis using Microsoft Excel. Jeffrey Simonoff, (2002) at http://www.stern.nyu.edu/~jsimonof/classes/1305/pdf/excelreg.pdf

Bell C.E. Excel Statistical Functions. ***Proceedings of the Annual Meeting of the American Statistical Association*****, August 5-9, 2001**

(1) Knusel, L., On the accuracy of statistical distributions in Microsoft Excel 97, Computational Statistics and Data Analysis, 26, 375-377, 1998.

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

(3) Sawitzki, G., Report on the reliability of data analysis systems, Computational Statistics and Data Analysis, 18, 289-301, 1994.

Cryer J.D. Problems With Using Microsoft Excel for Statistics. ***Proceedings of the Annual Meeting of the American Statistical Association*****, August 5-9, 2001**

(1) Allen, I. E. (1999), "The Role of Excel for Statistical Analysis", Making Statistics More Effective in Schools of Business 14th Annual Conference Proceedings, ed. A. Rao, Wellesley: http://weatherhead.cwru.edu/msmesb/

(2) Callaert, H. (1999), "Spreadsheets and Statistics: The Formulas and the Words", *Chance*, 12, 2, p. 64.

(3) Cleveland, W. S., *Visualizing Data*, 1993, Hobart Press, Summit, NJ

(4) Cleveland, W. S., *The Elements of Graphing Data*, Revised Edition, 1994, Hobart Press, Summit, NJ

(5) Goldwater, Eva, Data Analysis Group, Academic Computing, University of Massachusetts, *Using Excel for Statistical Data Analysis: Successes and Cautions*, November 5, 1999, www-unix.oit.umass.edu/~evagold/excel.html

E Goldwater. Using Excel for Statistical Data Analysis, Data Analysis Group, Academic Computing, U Mass, 1997; updated 1999.

McCullough, B.D. and Wilson, B. (2002), "On the accuracy of statistical procedures in Microsoft Excel 2000 and Excel XP," ***Computational Statistics and Data Analysis*****, 40, 713–721.**

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

selected references

(1) Knüsel, L., 1998. On the accuracy of statistical distributions in Microsoft Excel 97. Comput. Statist. Data Anal. 26, 375-377.

(2) Ling, R.F., 1974. Comparison of several algorithms for computing sample means and variances. J. Amer. Statist. Assoc. 69, 859-866.

(3) Longley, J.W., 1967. An appraisal of computer programs for the electronic computer from the point of view of the user. J. Amer. Statist. Assoc. 62, 819-841.

(4) Marsaglia, G., 1996. DIEHARD: A battery of tests of randomness. http://stat.fsu.edu/pub/diehard.

(5) McCullough, B.D., 1998. Asssessing the reliability of statistical software: Part I. The Amer. Statist. 52, 358-366.

(6) McCullough, B.D., 1999a. Asssessing the reliability of statistical software: part II. The Amer. Statist., forthcoming.

(7) Rogers, J., Filliben, J., Gill, L., Guthrie, W., Lagergren, E., Vangel, M., 1998. StRD: statistical reference datasets for assessing the numerical accuracy of statistical software. NIST TN# 1396, National Institute of Standards and Technology, USA.

(8) Sawitzki, G., 1994a. Testing numerical reliability of data analysis systems. Comput. Statist. Data Anal. 18, 269-286.

(9) Sawitzki, G., 1994b. Report on the reliability of data analysis systems. Comput. Statist. Data Anal. (SSN) 18, 289-301.

(10) B.D. McCullough, B. Wilson / Computational Statistics & Data Analysis 31 (1999) 27-37 37

(11) Wilkinson, L., 1994. Practical guidelines for testing statistical software. In: Dirschedl, P., Ostermann, R. (Eds.), Computational Statistics. Physica-Verlag, Berlin, pp. 111-124.

Statistical analysis using Microsoft Excel. Jeffrey Simonoff, (2002) at http://www.stern.nyu.edu/~jsimonof/classes/1305/pdf/excelreg.pdf

selected references

(1) Knüsel, L. (2002), "On the reliability of Microsoft Excel XP for statistical purposes," *Computational Statistics and Data Analysis*, **39**, 109–110.
(2) Knüsel, L. (2005), "On the accuracy of statistical distributions in Microsoft Excel 2003," *Computational Statistics and Data Analysis*, **48**, 445–449.
(3) McCullough, B.D. (2002), "Does Microsoft fix errors in Excel?" Proceedings of the 2001 *Joint Statistical Meetings*.
(4) McCullough, B.D. and Wilson, B. (2005), "On the accuracy of statistical procedures in Microsoft Excel 2003," *Computational Statistics and Data Analysis*, to appear.
(5) Pottel, H. (2001), "Statistical flaws in Excel," (www.mis.coventry.ac.uk/_nhunt/pottel.pdf).