



ELSEVIER

Computational Statistics & Data Analysis 40 (2002) 713–721

COMPUTATIONAL
STATISTICS
& DATA ANALYSIS

www.elsevier.com/locate/csda

On the accuracy of statistical procedures in Microsoft Excel 2000 and Excel XP

B. D. McCullough^{a,*}, Berry Wilson^b

^a*Department of Decision Sciences, LeBow College of Business, Drexel University, Philadelphia, PA 19104, USA*

^b*Finance and Graduate Economics Department, Lubin School of Business, Pace University, New York, NY 10038, USA*

Received 1 September 2001; received in revised form 1 February 2002

Abstract

The problems that rendered Excel 97 unfit for use as a statistical package have not been fixed in either Excel 2000 or Excel 2002 (also called “Excel 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. © 2002 Elsevier Science B.V. All rights reserved.

1. Introduction

McCullough (1998) proposed a methodology for assessing the reliability of statistical software in three areas: estimation, statistical distributions, and random number generation. Estimation is assessed using the “Statistical Reference Datasets” produced by the (American) National Institute of Standards and Technology,¹ which has four suites of tests: univariate summary statistics, one-way ANOVA, linear regression, and nonlinear least squares. Statistical distributions (e.g., for calculating p -values) are assessed using Knüsel’s (1989) ELV program. The random number generator (RNG) is subjected to the battery of randomness tests in Marsaglia (1996) DIEHARD program.

McCullough and Wilson (1999, “MW99”) applied this methodology to Excel 97, observed that Excel 97 was deficient in all three areas, and concluded that Excel should not be used for statistical analysis of data. Our conclusion has even been seconded by authors of statistical textbooks that teach students to use Excel, e.g., Sincich et al.

* Corresponding author.

E-mail addresses: bdmccullough@drexel.edu (B. D. McCullough), bwilson@pace.edu (B. Wilson).

¹ www.itl.nist.gov/div898/strd.

(2002, p. 51): “The use of Microsoft Excel in this text does not constitute an endorsement of its use in place of statistical packages in real-world applications”. See also Keller (2001, p. 14), Levine et al. (2002, p. 6) and Weiers (2002, p. xxxii), *inter alia*. Excel “add-ins” are not necessarily the answer, since these, too, can have flaws. Nerlove (2001) notes that the linear regression procedure in one statistical add-on for Excel “did not do significantly better than Excel[.]”. Because a comparison of Excel and a statistical package is of direct relevance, and also to increase the number of statistical packages for which StRD results have been published, an Appendix compares Excel and Stata v. 7 (Statacorp, 2001).

After the publication of MW99, both Excel 2000 and Excel XP were released. Each time we received numerous inquiries from concerned users of Excel, wondering whether it was safe to use the new version of Excel for statistical purposes. This is an important question, for at least two reasons. First, it is not inconceivable that more basic statistical calculations are performed in Excel than in all other statistical packages combined (Wilkinson, 1994, p. 114). Second, it is customary for software developers to use a new version or upgrade as an opportunity to fix known errors. Consequently, some users might assume that Microsoft would not release a new version without fixing known errors. Such an assumption, however, would prove incorrect.

2. Results

All the Excel 97 errors reported in MW99 exist in Excel 2000. Microsoft did slightly improve the quality of the inverse normal function in Excel XP, but even so its quality is still sub-par. This is discussed in Section 2.1. Additionally, there was an error with the standard normal RNG in Excel 97 that MW99 did not report. This error was changed, not fixed, in Excel 2000, and this new error was repeated in Excel XP. This is discussed in Section 2.2. Section 3 presents the conclusions.

2.1. Inverse normal function

Knüsel (1989) wrote the ELV program to compute statistical distributions with great accuracy. He used this program to assess the accuracy of the statistical distributions in Excel 97, and found many errors. These nontrivial errors affected many distributions, including: the Poisson, the binomial, the hypergeometric, the standard normal, the inverse standard normal, and others. His conclusion (Knüsel, 1998) was unambiguous: “So one has to warn statisticians against using Excel functions for scientific purposes”. All these same errors still exist in Excel 2000, and so Knüsel’s admonition applies to Excel 2000. With the sole exception of the inverse normal distribution, these same errors exist in Excel XP. Table 1 displays results for the inverse standard normal function. The ELV program is used to produce the “exact” result, while the Excel result is produced with $\text{NORMSINV}(x)$.

Researchers who use distributions for nothing more than hypothesis testing might not be alarmed by these discrepancies. Consider, however, trying to do “Six Sigma” calculations (e.g., Breyfogle, 1999) with Excel 97/2000, as opposed to Excel XP.

Table 1
Inverse normal distribution

x	Exact	Excel		
		97/2000	XP	Stata
0.001	– 3.09023	– 3.09024	– 3.09025	– 3.09023
0.0001	– 3.71902	– 3.71947	– 3.71909	– 3.71902
1E-5	– 4.26489	– 4.26546	– 4.26504	– 4.26489
1E-6	– 4.75342	– 4.76837	– 4.75367	– 4.75342
3E-7	– 4.99122	– 7.15256	– 4.99152	– 4.99122
2E-7	– 5.06896	– 5 000 000	– 5.06928	– 5.06896

Observe, though, that even after Microsoft “fixed” the NORMSINV function, Excel XP is accurate only to two decimals when $x = 1E-5$. One cannot help but wonder why Microsoft replaced a weak algorithm with a slightly less weak algorithm, instead of with a robust algorithm.

2.2. The standard normal RNG

MW99 noted that the Excel RNG fails randomness tests, its algorithm is not documented, and it is period deficient. These flaws are common to Excel 97, Excel 2000 and Excel XP. Rotz et al. (2002) conduct an even more extensive analysis of Excel’s RNG, and reach similar conclusions. Here we focus on Excel’s randomly generated standard normal deviates.

Excel 97 has a problem producing standard normal deviates. To see this, use the pulldown menus for “Tools”, “Data Analysis”, and “RNG” to generate some standard normal deviates. Set “Number of Variables” to 10, set “Number of Random Numbers” to 2000, and set “Random Seed” to 123. Then inspect cells 253:H, 1212:G, and 1245:D to observe what are allegedly standard normal deviates with a value of $-5\,000\,000$. Continue on to inspect cell 1845:F to find a value of $5\,000\,000$. These are not isolated occurrences: these values occur frequently in the generation of standard normal variates. We use the above seed only for purposes of demonstration; this flaw occurs for many seeds.

Users of Excel 2000 and Excel XP will find more sensible looking, but still incredibly unlikely values of -9.53674 and 5.364418 in place of $-5\,000\,000$ and $5\,000\,000$, respectively (all the other random numbers appear to be the same as in Excel 97). In no way can this change be considered a correction of the problem. Neither of these values should be seen in a sample of size 20 000, and a single occurrence of such an extreme number can ruin many types of simulations, including option pricing, inventory models, and extreme value analyses. In the same way that we can expect a standard normal deviate to fall outside ± 1.96 once every 20 calls, it should fall outside ± 5.364418 once every 12 million calls. A standard normal deviate should fall outside ± 9.53674 once every 678 000 000 000 000 000 000 calls, which amounts to making 1 billion calls per second for 20 000 years. Again setting the random seed to 123, and this time using Excel XP, in the first 20 000 calls we observe

allegedly standard normal variates with a value of -9.53674 three times, and 5.364418 one time.

Indeed, one must wonder why Microsoft made this particular change instead of fixing the problem correctly. The use of $\pm 5\,000\,000$ at least has the virtue that it is more likely to produce nonsensical simulation results, and thus be caught by a user before he relies on erroneous results. By contrast, the use of -9.53674 and 5.364418 can result in an incorrect simulation answer, but one that is not so likely to be caught by the user. To wit: what was once a serious but possibly noticeable error has become one of what the eminent computer scientist Kahan (2000) calls “errors of the worst kind: too small to be obvious but too big to be tolerable, ... too rare to be discovered by the customary desultory testing”.

3. Conclusions

Using Wilkinson (1985) *Statistics Quiz*, Sawitzki (1994a, b) noted flaws in Excel 4.0, including the unstable calculation of the sample variance. Microsoft has since taken Excel through five major revisions (5.0, Excel 95, Excel 97, Excel 2000, and Excel XP) without correcting even this easily remedied flaw. One has to wonder whether Microsoft really considers this, and other flaws, to be problems that need solving—see McCullough (2002) for further discussion of this point. This is especially important given that the only two “fixes” we could observe—to the standard normal RNG and the inverse normal distribution—did not raise the quality of these procedures to acceptable levels. Further doubt on the numerical reliability of Microsoft Excel is cast by a recent report from the United Kingdom’s National Physical Laboratory (Cook et al., 1999), which is available on the web. Also of interest to statistical users of Microsoft Excel are the papers by Cryer (2002) and Carlson (2002).

McCullough and Wilson (1999) examined Microsoft Excel 97 and concluded “Persons desiring to conduct statistical analyses of data are advised not to use Excel”. An examination of Excel 2000 and Excel XP provides no reason to revise that conclusion.

Acknowledgements

A previous version of this paper was presented at a panel discussion on “The Use of Microsoft Excel for Statistical Analyses” at the 2001 ASA meetings in Atlanta. We thank other panel members and participants for valuable suggestions, as well as Malcolm Hazel and Dave Burmaster. We are especially grateful to an associate editor and two anonymous referees, who made many useful comments.

Appendix A

The programs for obtaining the Stata results are available at the Stata website.²

² <http://www.stata.com/support/cert/nist/index.html>.

Table 2

StRD results for univariate summary statistics. This table shows the number of accurate digits for \bar{x} , s and ρ (the mean, standard deviation and first-order autocorrelation coefficient). These results are the same for Excel 97, Excel 2000 and Excel 2002

Dataset	Excel			Stata		
	$\lambda_{\bar{x}}$	λ_s	λ_ρ	$\lambda_{\bar{x}}$	λ_s	λ_ρ
Pidigits (l)	15	15	4.0	15	15	14.9
Lottery (l)	15	15	2.1	15	15	15
Lew (l)	15	15	2.6	15	15	14.8
Mavro (l)	15	9.4	1.8	15	13.1	13.7
Michelso (l)	15	8.3	3.6	15	13.8	13.4
Numacc1 (l)	15	15	0	15	15	15
Numacc2 (a)	14.0	11.6	3.3	15	15	15
Numacc3 (a)	15.0	1.1	3.3	15	9.5	11.9
Numacc4 (h)	14.0	0	3.3	15	8.3	10.7

Each of the four suites of StRD tests contains several problems of varying degree of difficulty: low (l), average (a), and high (h). For each problem, NIST computed the correct answer, say ‘ c ’, to several digits (15 digits for linear problems, 11 digits for nonlinear problems). For an answer produced by a statistical package, say, ‘ x ’, the number of correct digits can be calculated via the *log relative error* as

$$\lambda = \text{LRE}(x) = -\log_{10} \left(\frac{|x - c|}{|c|} \right).$$

For example, if $c = 6.54321$ and $x = 6.54300$, then $\text{LRE}(x) = 4.5$ correct digits.

A.1. Univariate summary statistics

This suite of tests has benchmark values for the mean (\bar{x}), the sample standard deviation (s), and the first-order autocorrelation coefficient (ρ). The Excel commands for computing these quantities are: ‘average’, ‘stdev’ and ‘correl’. Results are presented in Table 2. It can be deduced that *Excel uses unstable algorithms for calculation of the sample standard deviation and for calculation of the correlation coefficient*. Excel’s performance on this suite of tests is unacceptable.

A.2. Analysis of variance

Since ANOVA produces many numerical results, here only the LRE for the final F -statistic is presented. Results are presented in Table 3. The negative sums-of-squares produced by Excel implies that *Excel uses an unstable algorithm for one-way ANOVA calculations*. Excel’s performance on this suite of tests is unacceptable.

The Stata results for the higher difficulty problems are the best that can be achieved using conventional algorithms and double precision calculation on a PC (McCullough, 2000a). It should be noted that the package *Mathematica* can surpass these

Table 3
StRD results for ANOVA. This table shows the number of accurate digits in the final “F”-statistic. These results are the same for Excel 97, Excel 2000 and Excel 2002

Dataset	Excel	Stata
SiResist (l)	8.5	13.1
Simon1 (l)	14.3	14.4
Simon2 (l)	12.5	13.3
Simon3 (l)	12.6	14.7
Simon4 (l)	1.7	10.4
AgWt (a)	1.8	10.2
Simon5 (a)	1.1	10.2
Simon6 (a)	0 ^a	10.2
Simon7 (h)	0 ^b	4.4
Simon8 (h)	0 ^a	4.4
Simon9 (h)	0 ^a	4.2

^aNegative within group sum of squares.

^bNegative between group sum of squares.

Table 4
StRD linear regression results. This table shows the number of accurate digits for the least accurate coefficient ($\hat{\beta}$) and the least accurate standard error thereof ($\hat{\sigma}$). These results are the same for Excel 97, Excel 2000 and Excel 2002

Dataset	Excel		Stata	
	$\lambda_{\hat{\beta}}$	$\lambda_{\hat{\sigma}}$	$\lambda_{\hat{\beta}}$	$\lambda_{\hat{\sigma}}$
Norris (l)	12.1	13.8	12.8	13.5
Pontius (l)	11.2	14.3	11.5	13.0
Origin1 (a)	14.7	15	14.7	15
Origin2 (a)	15	15	15	15
Filip (h)	0	0	No solution	
Longley (h)	7.4	8.6	12.1	12.9
Wampler1 (h)	6.6 ^a	7.2	6.9	15
Wampler2 (h)	9.7	11.8	9.7	15
Wampler3 (h)	6.6	11.2	6.5	10.8
Wampler4 (h)	6.6	11.2	6.5	10.8
Wampler5 (h)	6.6	11.2	6.4	10.8

^aIncorrectly reported as 7.0 in MW99 Table 6.

conventional limits, and return a perfect score on all aspects of the StRD (McCullough, 2000b).

A.3. Linear regression

Since linear regression produces many numerical results, here only the lowest LRE for all the estimated coefficients ($\hat{\beta}$) and the lowest LRE for their standard errors ($\hat{\sigma}$) are presented. Results are presented in Table 4.

A problem that is “too collinear” for the solver to handle can result in a completely inaccurate answer, due to the accumulation of rounding error. Therefore, it is important for the package to determine whether the answer is likely to be completely corrupted by cumulated rounding error and if so not to display the “answer”. This is a very important aspect of a general purpose linear regression routine (Press et al., 1992, p. 23).

An easy way to probe this important aspect of a linear regression procedure is to give it a nearly singular problem, e.g., the test problem Filip is a tenth-order polynomial. This is just an easy way to see whether the package checks for near-singularity, and is unrelated to whether users actually run such regressions in practice.

Stata checks the data, determines that the problem is too collinear for the solver to produce an accurate solution, and returns an error message to this effect. This is a reliable solution, for the package has not misled the user. Excel, in contrast, produces a “solution” that is completely corrupted by cumulated rounding error. The relevant point is that *the Excel linear regression procedure either does not check the data for excessive collinearity, or does a bad job of checking*, and on this basis Excel’s linear regression procedure can be judged unacceptable.

A.4. Nonlinear regression

The nonlinear test problems come with two sets of starting values: Start I is “far” from the solution and makes the problem more difficult to solve; Start II is “close” to the solution and makes the problem easier to solve. As in the case of linear regression, if the package cannot produce an accurate solution, the package should recognize this situation and report the fact to the user. All test problems are run from Start I; Start II is used only in the case that the package informs the user that it cannot produce a solution from Start I.

Nonlinear solvers frequently offer a variety of options, and one use of the StRD is to determine the set of options that produces optimal performance. Here we use the options determined by McCullough and Wilson (1999): convergence tolerance = $1E-7$ with automatic scaling. Again we present the least accurate of the coefficients; the Excel Solver does not produce standard errors. Results are presented in Table 5.

From Start I, Stata cannot find accurate solutions to the test problems MGH09, MGH10, Eckerle4 and Rat43. However, Stata informs the user that solutions cannot be found to these problems; and Stata finds correct solution from Start II. By contrast, Excel never fails to produce a solution from Start I, though these “solutions” often are accurate to zero digits.

We did find a slight, but negligible improvement in Excel XP, which returned a few results differing from those in Table 5: Misra1a = 5.0, Misra1b = 5.8, Nelson = 1.1.

The conclusion concerning the use of Excel for solving these types of problems is the same for Excel 2002 as for Excel 2000 and Excel 97: *the Solver has a marked tendency to stop at a point that is not a solution and declare that it has found a solution*, and therefore Excel’s performance on this suite of tests is unacceptable.

Table 5

StRD nonlinear results. This table shows the number of accurate digits in the least accurate of the coefficients. These results are the same for Excel 97 and Excel 2000

Dataset	Excel	Stata
Misra1a (l)	4.8	9.1
Chwirut2 (l)	4.6	7.9
Chwirut1 (l)	4.9	7.6
Lanczos3 (l)	0	6.2
Gauss1 (l)	0	8.6
Gauss2 (l)	0	8.2
DanWood (l)	5.5	8.6
Misra1b (l)	4.4	8.3
Kirby2 (a)	1.1	9.1
Hahn1 (a)	0	7.1
Nelson (a)	1.3	7.1
MGH17 (a)	0	9.4 ^a
Lanczos1 (a)	0	10.6
Lanczos2 (a)	0	7.4
Gauss3 (a)	0	8.2
Misra1c (a)	4.6	9.2
Misra1d (a)	5.3	9.3
Roszman1 (a)	3.7	7.9
ENSO (a)	3.4	4.7
MGH09 (h)	0	7.0 ^a
Thurber (h)	1.8	6.5
BoxBod (h)	0	7.3
Rat42 (h)	5.3	7.6
MGH10 (h)	0	7.5 ^a
Eckerle4 (h)	0	8.3 ^a
Rat43 (h)	0	6.0 ^a
Bennett5 (h)	0	6.3

^aDenotes a solution from Start II.

References

- Breyfogle, F., 1999. *Implementing Six Sigma: Smarter Solutions Using Statistical Methods*. Wiley, New York.
- Carlson, W., 2002. The use of microsoft excel for statistical purposes: a textbook author perspective. In: *Proceedings of the 2001 Joint Statistical Meetings [CD-ROM]*, American Statistical Association, Alexandria, VA.
- Cook, H.R., Cox, M.G., Dainton, M.P., Harris, P.M. 1999. Testing the intrinsic functions of Excel. NPL Report CISE 27/00, www.npl.co.uk/ssfm/download.
- Cryer, J., 2002. Problems with using Microsoft Excel for statistics. In: *Proceedings of the 2001 Joint Statistical Meetings [CD-ROM]*, American Statistical Association, Alexandria, VA.
- Kahan, W., 2000. Marketing versus mathematics. Manuscript, EE & CS Department, University of California, Berkeley.
- Keller, G., 2001. *Applied Statistics with Microsoft Excel*. Duxbury, Belmont, CA.
- Knüsel, L., 1989. *Computergestützte berechnung statistischer verteilungen*. Oldenburg, München-Wien (An English version of the program can be obtained at www.stat.uni-muenchen.de/~knuesel/elv).

- Knüsel, L., 1998. On the accuracy of statistical distributions in Microsoft Excel 97. *Comput. Statist. Data Anal.* 26, 375–377.
- Levine, D., Stephan, D., Krehbiel, T., Berenson, M., 2002. *Statistics for Managers Using Microsoft Excel*, 3rd Edition. Prentice-Hall, Englewood Cliffs, NJ.
- Marsaglia, G., 1996. DIEHARD: a battery of tests of randomness. <http://stat.fsu.edu/~geo>.
- McCullough, B.D., 1998. Assessing the reliability of statistical software: Part I. *Amer. Statist.* 52, 358–366.
- McCullough, B.D., 2000a. Experience with the StRD: application and interpretation. In: *Statistics and Computing. Proceedings of the Interface*, Vol. 31, pp. 16–21.
- McCullough, B.D., 2000b. The accuracy of Mathematica 4 as a statistical package. *Comput. Statist.* 15, 279–299.
- McCullough, B.D., 2002. Does Microsoft fix errors in Excel? In: *Proceedings of the 2001 Joint Statistical Meetings [CD-ROM]*, American Statistical Association, Alexandria, VA.
- McCullough, B.D., Wilson, B., 1999. On the accuracy of statistical procedures in Microsoft Excel 97. *Comput. Statist. Data Anal.* 31, 27–37.
- Nerlove, M., 2001. On the numerical accuracy of Mathematica 4.1 for doing ordinary least squares regression. Manuscript, AREC, University of MD.
- Press, W., Teukolsky, S., Vetterling, W., Flannery, B., 1992. *Numerical Recipes in Fortran*, 2nd Edition. Cambridge University Press, New York.
- Rotz, W., Falk, E., Wood, D., Mulrow, J., 2002. A comparison of random number generators used in business. In: *Proceedings of the 2001 Joint Statistical Meetings [CD-ROM]*, American Statistical Association, Alexandria, VA.
- Sawitzki, G., 1994a. Testing the numerical reliability of data analysis systems. *Comput. Statist. Data Anal.* 18, 269–286.
- Sawitzki, G., 1994b. Report on the numerical reliability of data analysis systems. *Comput. Statist. Data Anal. (Statist. Software Newslett.)* 18, 289–301.
- Sincich, T., Levine, D., Stephan, D., 2002. *Practical Statistics by Example Using Microsoft Excel*. Prentice-Hall, Upper Saddle River, NJ.
- Statacorp, 2001. *Stata Statistical Software: Release 7.0*. Stata Corporation, College Station, TX.
- Weiers, R., 2002. *Introduction to Business Statistics*, 4th Edition. Duxbury, Belmont, CA.
- Wilkinson, L., 1985. Statistics quiz: problems which reveal deficiencies in statistical programs. SYSTAT, Mimeo.
- Wilkinson, L., 1994. Practical guidelines for testing statistical software. In: Dirschedl, P., Ostermann, R. (Eds.), *Computational Statistics*. Physica-Verlag, Berlin, pp. 111–124.