Fixing Statistical Errors in Spreadsheet Software: The Cases of Gnumeric and Excel

B. D. McCullough*
Department of Decision Sciences
LeBow College of Business
Drexel University
Philadelphia, PA 19104
bdmccullough@drexel.edu

Abstract

The open source spreadsheet package "Gnumeric" was such a good clone of Microsoft Excel that it even had errors in its statistical functions similar to those in Excel's statistical functions. When apprised of the errors in v1.0.4, the developers of Gnumeric indicated that they would try to fix the errors. Indeed, Gnumeric v1.1.2, has largely fixed its flaws, while Microsoft has not fixed its errors through many successive versions. Persons who desire to use a spreadsheet package to peform statistical analyses are advised to use Gnumeric rather than Excel.

*Thanks to Berry Wilson for comments.

1 Introduction

It has long been known, at least informally, that Microsoft does not fix errors in the statistical procedures in Excel (McCullough, 2002). McCullough (1998) proposed a methodology to assess statistical and econometric software by examining three areas: statistical distributions, estimation, and random number generation. The present document is terse, and assumes that the reader already has at least read McCullough and Wilson (1999), which applies McCullough's (1998) methodology to Microsoft Excel 97.

This researcher first became acquainted with the Excel-clone "Gnumeric" in its 0.67 release. At that point in time, Gnumeric 1.0.4 already had been released. This researcher informally applied some of the usual tests (*e.g.*, those applied to Excel 97 in McCullough and Wilson (1999)) to Gnumeric, found it wanting, and informed the developers. The few part-time volunteers who maintain and develop Gnumeric fixed almost all the problems in a few weeks. The formal release of the fixed version was delayed for about a year, waiting for the release of a major revision (from 1.4 to 2.x) of a related package (Gnome).

This report compares improvements in the statistical functions of Gnumeric versions 0.67 and 1.1.2 to improvements in the statistical functioncs of Excel versions 97, 2000 and XP. It makes limited use (for the Poisson, Binomial and Inverse Beta distributions) to some results for Excel 2003 from the paper by Knüsel (2003).

2 Statistical Distributions

Accurate ("exact") values for statistical distributions, such as the normal distribution, are computed via Knüsel's (1989) ELV program and verified by comparison with SAS, as per Knüsel (1998, 2003). Knüsel (1998) described the desirable features of statistical functions. At a minimum, the function should be accurate to all digits that are displayed by default. If the package, by default, displays six digits but produces answers that are accurate only to two digits, the package is judged unacceptable. If a package produces "no result" for some inputs, this is noted. For all the values of the inputs used in these tests, algorithms are known that will provide acceptable answers. Consequently, to replace one defective algorithm with another is not evidence that the developers are familiar with customary practices in the field of statistical computing.

This section is not a comprehensive comparison, as not all the functions in Gnumeric were tested. The comparison is merely illustrative. The specific distributions chosen were those highlighted by Knüsel (1998 and 2003) in his analyses of Excel.

As can be seen in Table 1, Excel's Poisson distribution returned no result for values near the mean of the distribution in old versions. In Excel 2003, Microsoft obtained an accurate answer in the central region of the distribution in exchange for inaccurate results in the tail. This is not a good "fix". A good fix is demonstrated by Gnumeric, where an algorithm that was inaccurate in the central region and the tail was exchanged for an algorithm that provides exact results in both areas.

		Excel		gnumeric	
k	exact	97/2000/XP	2003	v0.67	v1.1.2
0	1.3839E-87	exact	0	no result	exact
10	4.1096E-71	exact	0	0	exact
50	6.8158E-37	exact	0	0	exact
100	3.723 64 E-15	exact	0	3.77476E-15	exact
103	2.8916 E-14	exact	0	2.86658E-14	exact
104	5.6170 E-14	exact	2.7254 E-14	5.61773E-14	exact
110	2.4813 E-12	exact	2.4524 E-12	2.48124E-12	exact
133	2.943 90 E-07	exact	exact	exact	exact
134	4.456 17 E-07	no result	exact	exact	exact
200	0.518 795	no result	exact	5.18794E-01	exact
250	0.999 715	no result	exact	exact	exact

Table 1: Poisson Distribution with $\lambda = 200$, $P(X \leq k)$

As can be seen in Table 2, the hypergeometric distribution in Gnumeric v0.67 returned zeroes for values that were not close to zero, but this was fixed in Gnumeric v1.1.2. Microsoft did not fix Excel.

		Excel	Gnumeric	
k	exact	97/2000/XP	v0.67	v1.1.2
225	3.865 27 E-4	no result	0.0	exact
250	0.049 7072	no result	0.0	exact
275	3.865 27 E-4	no result	0.0	exact

Table 2: Hypergeometric Distribution, $N = 1030, M = 515, n = 500$, $P(X = k)$

As can be seen in Table 3, Microsoft again obtained an accurate answer in the central region of the distribution at the cost of providing inaccurate answers in the tail. Gnumeric displayed no problem that needed fixing.

		Excel		Gnumeric	
k	exact	97/2000/XP	2003	v0.67	v1.1.2
10	3.111E-287	exact	0	exact	exact
50	3.941E-225	exact	0	exact	exact
100	1.394E-169	exact	0	exact	exact
200	5.4578E-92	exact	0	exact	exact
390	3.1820E-15	exact	0	exact	exact
391	5.2410E-15	exact	2.0590E-15	exact	exact
400	3.897 35 E-13	exact	3.8655E-13	exact	exact
499	0.167 042	exact	exact	exact	exact
500	0.183 106	no result	exact	exact	exact
515	0.512 428	no result	exact	exact	exact
550	0.986 550	no result	exact	exact	exact
575	0.999 920	no result	exact	exact	exact

Table 3: Binomial Distribution with $n = 1030$ and $p = 0.5$, $P(X \leq k)$

As can be seen in Table 4, Gnumeric's standard normal distribution displayed no problem that needed fixing. Microsoft did not fix Excel.

			Gnumeric	
x	exact	Excel 97/2000/XP	v0.67	v1.1.2
-3	0.001 349 90	0.001 349 967	exact	exact
-4	3.167 12 E-5	3.168 60 E-5	exact	exact
-5	2.866 52 E-7	2.871 05 E-7	exact	exact
-6	9.865 88 E-10	9.901 22 E-10	exact	exact
-8.2	1.201 94 E-16	1.110 22 E-16	exact	exact
-8.3	5.205 57 E-17	0	exact	exact

Table 4: Standard Normal Distribution, $P(X < x)$

As can be seen in Table 5, Gnumeric's inverse normal distribution displayed no problem that needed fixing. In the XP release, Microsoft attempted to fix Excel, but did not do a very good job. Microsoft traded a weak algorithm for one that was slightly less weak, instead of using a robust algorithm.

		Excel		Gnumeric	
x	exact	97/2000	XP	v0.67	v1.1.2
0.001	-3.09023	-3.09024	-3.09025	exact	exact
0.0001	-3.71902	-3.71947	-3.71909	exact	exact
1E-5	-4.26489	-4.26546	-4.26504	exact	exact
1E-6	-4.75342	-4.76837	-4.75367	exact	exact
3E-7	-4.99122	-7.15256	-4.99152	exact	exact
2E-7	-5.06896	-5000000	-5.06928	exact	exact

Table 5: Inverse Normal Distribution, $P(X < x) = p$

As can be seen in Table 6, the inverse chi square distribution Gnumeric v0.67 was a weak, but this was fixed in Gnumeric v.1.1.2. Microsoft did not fix Excel.

			Excel	Gnumeric	
p	n	exact	97/2000/XP	v0.67	v1.1.2
0.001	1	10.8276	10.827359 88	exact	exact
1e-6	1	23.9281	24.36637878	no result	exact
0.001	10	29.5883	29.58788536	exact	exact
1e-6	10	46.8630	46.7658625	no result	exact

Table 6: Inverse Chisquare, $P(X < x) = p$

As can be seen in Table 7, Gnumeric's inverse F distribution displayed no problem that needed fixing. Microsoft did not fix Excel.

			Excel	Gnumeric	
p	$n_1 = n_2$	exact	97/2000/XP	v0.67	v1.1.2
0.001	2	999	998.8434613	exact	exact
1e-6	2	999999	976562.5	exact	exact
0.001	5	29.7524	29.75139068	exact	exact
1e-6	5	492.881	476.8371582	exact	exact
0.001	10	8.753 87	8.753886505	exact	exact
1e-6	10	40.0156	40.97819328	exact	exact

Table 7: Inverse F, $P(X < x) = p$

As can be seen in Table 8, Gnumeric's inverse t distribution displayed no problem that needed fixing. Microsoft did not fix Excel.

p	n	exact	Excel 97/2000/XP	Gnumeric v0.67	Gnumeric v1.1.2
0.001	2	31.5991	31.599 774 96	exact	exact
1e-6	2	999.999	915.527 3438	exact	exact
0.001	5	6.868 83	6.868 503 988	exact	exact
1e-6	5	28.4785	28.610 229 49	exact	exact
0.001	10	4.586 89	4.586 763 68	exact	exact
1e-6	10	10.5165	10.728 836 06	exact	exact

Table 8: Inverse t, $P(X < x) = p$

As can be seen in Table 9, Excel produces the same answers for all four versions. This is surprising, because Microsoft claims to have fixed the inverse beta function. Gnumeric exchanged a weak algorithm that produced no results for better algorithm.

p	α	β	exact	Excel 97/2000/XP/2003	Gnumeric v0.67	Gnumeric v1.1.2
0.001	5	2	0.181 386	0.181 396	no result	exact
1E-6	5	2	0.044 427	0.042 969	no result	exact
0.001	10	100	0.027 946	exact	no result	exact
1e-6	10	100	0.012 149	0.011 719	no result	exact

Table 9: Inverse β, $P(X < x) = p$

3 StRD

The "Statistical Reference Datasets" (StRD) (www.nist.gov/itl/div898/strd) comprises four suites of benchmark tests for statistical software: univariate summary statistics, analysis of variances (ANOVA), linear regression, and nonlinear regression. The nonlinear regression suite is not included in this comparison because Gnumeric does not offer this procedure.

Each of the 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

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

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

3.1 StRD: Univariate Summary Statistics

This suite of tests has benchmark values for the mean ($\bar{x}$) and the sample standard deviation (s). As can been seen, Gnumeric 0.67 used an unstable algorithm for computing the sample standard deviation, and on this basis its performance can be considered unacceptable. This was fixed in Gnumeric 1.1.2. Microsoft did not fix the identical problem in Excel.

	Excel 97/2000/XP		Gnumeric v0.67		Gnumeric v1.1.2	
dataset	$\lambda_{\bar{x}}$	λ_s	$\lambda_{\bar{x}}$	λ_s	$\lambda_{\bar{x}}$	λ_s
Pidigits (l)	15	15	15	15	15	14.9
Lottery (l)	15	15	15	15	15	15
Lew (l)	15	15	15	15	15	15
Mavro (l)	15	9.4	15	13.1	15	13.1
Michelso (l)	15	8.3	15	13.9	15	13.9
Numacc1 (l)	15	15	15	15	15	15
Numacc2 (a)	14.0	11.6	15	15	15	14.2
Numacc3 (a)	15.0	1.1	15	9.5	15	9.5
Numacc4 (h)	14.0	0	15	8.3	15	8.3

Table 10: StRD results for univariate summary statistics.

3.2 StRD: ANOVA

Since ANOVA produces many numerical results, here only the LRE for the final F-statistic is presented. Results are presented in Table 11. The negative sums-of-squares produced by Excel implies that Excel uses an unstable algorithm. Excel's performance on this suite of tests is unacceptable. Gnumeric 0.67 used a good algorithm (a good algorithm will fail the very demanding Simon9 test) but switched to an even better algorithm for v1.1.2. See McCullough (2000) for a discussion of the ability of various algorithms to solve the Simon9 problem.

dataset	Excel 97/2000/XP	Gnumeric v0.67	Gnumeric v1.1.2	dataset	Excel 97/2000/XP	Gnumeric v0.67	Gnumeric v1.1.2
SiResist (l)	8.5	12.4	12.7	Simon5 (a)	1.1	8.3	9.3
Simon1 (l)	14.3	14.3	15.0	Simon6 (a)	0^a	6.5	9.2
Simon2 (l)	12.5	14.3	14.9	Simon7 (h)	0^b	2.7	3.3
Simon3 (l)	12.6	13.4	13.8	Simon8 (h)	0^a	2.2	3.3
Simon4 (l)	1.7	8.5	9.3	Simon9 (h)	0^a	0	3.2
AgWt (a)	1.8	8.5	8.8				

a – negative within group sum of squares
b – negative between group sum of squares

Table 11: StRD results for ANOVA.

3.3 StRD: 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 12.

To return zero digits of accuracy for a regression problem indicates that the package's procedure either doesn't check for near singularity of the design matrix, or does a very bad job of checking. Either way, returning zero digits of accuracy is unacceptable. Gnumeric 0.67 returns zero digits for the Filip problem, and this is unacceptable. Gnumeric 0.67 refuses to give a solution to the Longley and Wampler problems, and specifically says that the design matrix is near singular. This is an acceptable response, because the user is not misled. Gnumeric 1.1.2 was fixed. Microsoft did not fix Excel.

4 Random Number Generator

Ripley (1990) describes the important characteristics of an RNG that is to be used for statistical purposes: (1) it should be reproducible from a simply specified starting point; (2) it should have a very long period; (3) it should produce numbers that are very close to uniform; and (4) it should produce numbers that are very close to independent in a moderate number of dimensions.

The Gnumeric RNG is based on "/dev/urandom" which is a source of randomness based on environmental noise from device drivers and other operating system sources. This is a very good method to obtain seeds for pseudo-random number generators, but it is not a good way to generate random numbers for statistical purposes, for its sequences are not reproducible. Consequently, it cannot support the replication of experimentation that is critical to Monte Carlo studies, etc.

On this basis alone, the RNG in Gnumeric can be judged unacceptable for statis-

	Excel 97/2000/XP		Gnumeric v0.67		Gnumeric v1.1.2	
dataset	$\lambda_{\hat{\beta}}$	$\lambda_{\hat{\sigma}}$	$\lambda_{\hat{\beta}}$	$\lambda_{\hat{\sigma}}$	$\lambda_{\hat{\beta}}$	$\lambda_{\hat{\sigma}}$
Norris (l)	12.1	13.8	12.7	0	12.7	13.9
Pontius (l)	11.2	14.3	0	0	11.5	13.9
Origin1 (a)	14.7	15	15	15	14.7	15
Origin2 (a)	15	15	15	15	15	15
Filip (h)	0	0	0	0	no solution	
Longley (h)	7.4	8.6	no solution		8.1	9.7
Wampler1 (h)	6.6	7.2	no solution		7.4	8.0
Wampler2 (h)	9.7	11.8	no solution		10.5	12.7
Wampler3 (h)	6.6	11.2	no solution		7.4	11.8
Wampler4 (h)	6.6	11.2	no solution		7.4	11.8
Wampler5 (h)	6.6	11.2	no solution		7.4	11.8

Table 12: StRD linear regression results.

tical purposes. There is no reason to apply tests to the random numbers produced by Gnumeric. The random number generator in Excel has been examined by Rotz *et al.* (2002), and found to fail Marsaglia's (1996) DIEHARD tests.

5 Conclusions

Microsoft has repeatedly released new versions of Excel without correcting errors in its statistical procedures. As of this writing, Excel 2003 has been released, but Microsoft's claims of enhanced accuracy have yet to be independently verified. Preliminary results are available for the statistical distributions in Excel 2003, and these have been adduced. Four examples of Microsoft "fixing" a bug by replacing one inaccurate algorithm with another have been given (Poisson, Binomial and Inverse Beta in Excel 2003, and Inverse Normal Distribution in Excel XP). Meanwhile, the developers of Gnumeric, who faced many of the same problems, managed to fix their package quickly and correctly.

One might have wondered whether there is something inherently difficult about fixing statistical bugs in spreadsheet package. It is reasonable for the developer of a spreadsheet package not to fix errors in a timely fashion? The answer to this question is seen to be, No. A related question is whether amateurs working on a volunteer basis, with no budget for R&D, could do a better job than professionals (who work for pay and have an R&D budget) when it comes to fixing statistical errors in a spreadsheet. The answer to this question is, Yes.

It has been shown that the RNG in Gnumeric is not suitable for statistical purposes, because it does not generate reproducible sequences. This will not be for long, as the developers already have added the Mersenne Twister (Matsumoto and Nishimura, 1998) to a beta version of Gnumeric. Not only does the Mersenne Twister pass all of Marsaglia's (1996) DIEHARD tests, McCullough (2003) has shown that it also passes all of the more stringent tests in L'Ecuyer's (2003) TESTU01 suites of tests.

Persons who desire to use a spreadsheet package to peform statistical analyses are advised to use Gnumeric rather than Excel, at least until Microsoft's fixes to Excel 2003 have been shown to make Excel 2003 more reliable than Gnumeric.

References

Knüsel, Leo (1989), "Computergestütze Berechnung Statisticher Vereilungen, (ELV)" Oldenburg, München-Wien

Knüsel, Leo (1998), "On the Accuracy of Statistical Distributions in Microsoft Excel 97," *Computational Statistics and Data Analysis* **26**, 375-377

Knüsel, Leo (2003), "On the Accuracy of Statistical Distributions in Microsoft Excel 2003," manuscript

L'Ecuyer, P. (2003), "TESTU01: A Software Library in ANSI C for Empirical Testing of Random Number Generators"

Marsagliam, G. (1996), *The Marsaglia random number CDROM including the diehard battery of tests of randomness* http://stat.fsu.edu/pub/diehard/

Matsumoto, M. and T. Nishimura (1998), "Mersenne Twister: A 623-dimensionally Equidistributed Uniform Pseudo-random Number Generator," *ACM Transactions on Modeling and Computer Simulation* **8**, 3-30

McCullough, B. D. (1998), "Assessing the Reliability of Statistical Software: Part I," *American Statistician* **52**, 358-366

McCullough, B. D. (2000), "Experience with the StRD: Application and Interpretation," Computing Science and Statistics **31**, 16-21, 2000

McCullough, B. D. (2002), "Does Microsoft Fix Errors in Excel?" *Proceedings of the 2001 Joint Statistical Meetings* [CD-ROM], to appear

McCullough, B. D. (2003), "A Review of TESTU01," manuscript

McCullough, B. D. and Berry Wilson (1999), "On the Accuracy of Statistical Procedures in Microsoft Excel 97," *Computational Statistics and Data Analysis* **31**, 27-37

Ripley, Brian D. (1990), "Thoughts on Pseudorandom Number Generators," *Journal of Computational and Applied Mathematics* **31**(1), 153-163

Rotz, Wendy, Eric Falk, Daniel Wood, and Jeri Mulrow (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