

## On the accuracy of statistical procedures in Microsoft Excel 2010

Guy Mélard

Received: date / Accepted: date

**Abstract** All previous versions of Microsoft Excel until Excel 2007 have been criticized by statisticians for several reasons, including the accuracy of statistical functions, the properties of the random number generator, the quality of statistical add-ins, the weakness of the Solver for nonlinear regression, and the data graphical representation. Until recently Microsoft did not make an attempt to fix all the errors in Excel and was still marketing a product that contained known errors. We provide an update of these studies given the recent release of Excel 2010 and we have added OpenOffice.org Calc 3.3 and Gnumeric 1.10.16 to the analysis, for the purpose of comparison. The conclusion is that the stream of papers, mainly in Computational Statistics and Data Analysis, has started to pay off: Microsoft has partially improved the statistical aspects of Excel, essentially the statistical functions and the random number generator.

**Keywords** statistical function · statistical procedure · random number generator · nonlinear regression · Excel 2010 · OpenOffice.org Calc 3.3 · Gnumeric

---

Guy Mélard \*  
ECARES and Solvay Brussels School of Economics and Management, Université libre de Bruxelles, CP114/4 Avenue Franklin Roosevelt 50, B-1050 Bruxelles, Belgium.  
Tel: +3226504604 or +32476372514  
Fax: +3226504012  
E-mail: gmelard@ulb.ac.be

## 1 Introduction

### 1.1 Generalities

Criticisms on the statistical aspects of the various versions of Microsoft Excel started with Sawitzki (1994) and Knüsel (1998) for versions 4 and 6, respectively, and have been followed by several published papers, e.g. McCullough and Wilson (1999), McCullough and Wilson (2002), McCullough and Wilson (2005).

For a long time Microsoft ignored these criticisms from statisticians. Failure to address the problems seriously seems to have exasperated the statistical community. The first (timid) improvements appeared in Excel 2002 (included in Office XP) and Excel 2003 (included in Office 2003). In Excel 2000, the inverse normal distribution for probability  $2 \cdot 10^{-7}$  was  $-5000000$  (McCullough and Wilson, 2002) and was “corrected” to  $-5.06928$  in Excel 2002, instead of  $-5.06896$ . Excel 2003 came with a better handling of extreme observations, more robust statistical functions, and a different pseudo-random number generator. It is a general pattern that Microsoft, in its attempt to fix errors, actually introduces new errors. Other spreadsheet processors do exist, including the open source software packages OpenOffice.org Calc and Gnumeric but Microsoft Excel is still leading the market by a wide margin (Hümmer, 2010).

Recent contributions to the subject make up a whole section of Computational Statistics and Data Analysis published by McCullough (2008a), in particular Yalta (2008). See also Almiron et al. (2010) and Hargreaves and McWilliams (2010), as well as a very complete web site held by Heiser (2009) (where the history of Microsoft “fixes” is nicely described). In those papers and that web site, the latest versions covered are Excel 2007 and Calc 3.0. Meanwhile, Microsoft Office 2010 and OpenOffice 3.3 have been released, calling for an update of those studies. Note that Almiron et al. (2010) also covers implementations over several operating systems (Microsoft Windows, Linux Ubuntu, Apple MacOS X) and several hardware platforms (Intel i386 and AMD amd64) and covers not only several releases of Excel and Calc but also an OpenOffice derivative for MacOS, NeoOffice, and GNU Oleo for Linux Ubuntu.

Here are the main subjects of dislike by statisticians before the release of Excel 2010:

1. Accuracy of the statistical functions. Yalta (2008) has shown that Excel 2007 is identical to Excel 2003 for most functions but that Calc 3.0 often obtains more precise values, leaving room for improvement in Excel. He also notes that Gnumeric always obtains correct results. See also Knüsel (2005) and Almiron et al. (2010).
2. The generator of pseudo-random numbers. What is generally discussed is the `RAND()` function. See e.g. McCullough (2008b) and Almiron et al. (2010) for the basic criticisms. The latter paper also discusses OpenOffice.org Calc.

3. The quality of the statistical add-ins. The studies, e.g. McCullough and Heiser (2008), Almiron et al. (2010), have shown that several of the statistical add-ins are flawed, although the problems here are of different nature. External add-ins can also raise problems, see e.g. Yalta and Jenal (2009).
4. The Solver module. It is used in nonlinear fits. It often erroneously claims that convergence has been reached (McCullough and Heiser, 2008) and does not manage to solve many test problems, see McCullough and Wilson (1999), McCullough and Wilson (2002), Almiron et al. (2010).
5. The graphical representation of data. Basic default graphs are not good (Cryer, 2001) and there is no way to produce many of the statistical plots (especially histograms and boxplots) like in statistical packages, see e.g. Su (2008).

For quantitative results, accuracy is often measured by the number of correct significant digits of an estimate  $q$  with respect to a correct value  $c$ . It can be evaluated by the logarithm of relative error

$$\text{LRE} = \begin{cases} -\log_{10}(|q - c|/|c|), & \text{if } c \neq 0, \\ -\log_{10}(|q|), & \text{otherwise,} \end{cases} \quad (1)$$

and we set it to 15 if  $q = c = 0$ , or 16 if  $q = c \neq 0$ . A value of LRE less than 1 is set to 0, i.e. zero digit accuracy. Moreover, since the meaning of LRE is lost when  $q$  and  $c$  differ too much, it is often proposed (McCullough, 1998) to set LRE to 0 when they differ by a factor greater than 2, but that condition is rather vague and couldn't be implemented.

Most recent papers consider also Excel 2010: Knüsel (2011), Keeling and Pavur (2011), McCullough and Yalta (2013). Knüsel (2011) only deals with statistical functions of Excel 2010, mainly for extreme values of the arguments. Keeling and Pavur (2011) compare Excel 2007 and 2010 with Gnumeric 1.10.14, Google Docs (renamed Google Spreadsheet), Numbers 09, OpenOffice 3.3.0, Quattro Pro X4, R 2.13.0 and SAS 9.2 for the items 1, 2 and 3 in the list above. McCullough and Yalta (2013) compare three cloud spreadsheets including Excel Web App, sometimes with Excel 2007 and 2010. None of these papers present the results for statistical functions in terms of number of correct digits like we do. Since Keeling and Pavur (2011) made use of the Wilkinson's tests and the StRD datasets (see Section 4.1 for both), we have not duplicated their results. Another reason is given at the beginning of Section 4.2.

## 1.2 Excel 2010

Officially released on May 12, 2010, Office 2010 had pre-existed in beta version since the end of 2009. It preserves the ribbon of Office 2007 but with a new File menu (called "Backstage"). Here are some of its characteristics for statisticians:

- the accuracy of the functions, including statistical functions, is improved and new names appear in order to provide a more systematic terminology;
- the Solver add-in is said to have been improved;

- the graphs are improved: the number of points is increased with improved formatting and macro recording capability.

The object of the study is to assess the improvements of Excel 2010 in the stated areas. For more details about the improvements on Excel functions, see Microsoft (2009a). For the Solver module, we are only considering non-linear optimization without constraints since the other aspects would be out of the scope of this research devoted to statistics, not operational research. A fine analysis is necessary to check if the problems mentioned by McCullough and Wilson (2005) and Almiron et al. (2010) have been solved. On the other hand, we will examine OpenOffice.org Calc 3.3 and Gnumeric in parallel for the subjects 1-5 of Section 1.1. The intention is not to completely test these programs.

### 1.3 Calc 3.3

OpenOffice.org is an open source office suite that was launched in 2000 by Sun Microsystems to compete with Microsoft Office. It was based on the StarOffice suite purchased from the German company StarDivision. The suite is now distributed by Oracle which purchased Sun Microsystems in 2010. Let us note that some software companies, including IBM, Novell and Oracle themselves, have marketed custom versions of OpenOffice.org. Calc is the spreadsheet part of the suite. Currently (June 2011), version 3.3 is available. Calc 3.3 can read the .xlsx files created with Excel 2007 and Excel 2010 but Excel 2007 could not read the native .ods files created by Calc 3.3 before release of service pack 2. Both programs can however read Excel 2003 .xls files. In September 2010, some members of the OpenOffice.org Community Council decided to develop a LibreOffice suite outside of Oracle. In June 2011, Oracle donated the OpenOffice.org code to the Apache Software Foundation.

### 1.4 Gnumeric 1.10.16

It has been shown, notably in McCullough (2004c) and Almiron et al. (2010), that Gnumeric, an open source spreadsheet coming from the Linux Gnome community in the early 2000's, has high quality statistical functions. Therefore it is not necessary to repeat the detailed results here. McCullough (2004c, pg. 2) who had discovered some errors in a previous version has reported that "The few part-time volunteers who maintain and develop Gnumeric fixed all the problems in a few weeks."

**Table 1** Details from Yalta (2008) results with some additional cases and update to Excel 2010 and Calc 3.3 with respect to Mathematica 7 (Mma). Items like  $m/p + q$  in column “# cases” specify that Yalta (2008) tables contained  $m$  cases while original file had  $p$  cases and  $q$  additional cases were added. Column “min LRE” gives for Excel 2010 the minimum of LRE over all  $p + q$  cases in file. Weakest results are discussed in the separated subsections. Columns “Relative to Mma” give the maximum over  $p + q$  cases of the relative differences between Excel or Calc and Mathematica. Column “%Excel-Calc = 0” gives the percentage of differences between Excel and Calc results that are considered as 0 by Excel. See Sections 2.1 and 2.11 for comments. Note: # indicates several #value!

Excel function	Yalta table	#cases in table/file	min LRE	Relative to Mma		%Excel-Calc=0
				Excel	Calc	
BINOMDIST	2	14/19	12.8	1E-13	3E-15	42
HYPERGEOMDIST	3	10/15	13.1	6E-14	5E-15	0
POISSON	4	14/34	13.1	1E-14	1E-09	26
GAMMADIST	5	10/106	14.1	7E-15	8E-15	91
NORMINV	6	14/50	5.2	0	0	98
CHIINV	7	15/78+9	4.5	2E-05	1E-13	66
BETAINV	8	14/36+24	14.0	9E-15	9E-14	85
TINV #	9	15/72+8	5.4	3E-06	3E-06	80
FINV	10	12/72+24	11.0	8E-12	1E-11	57
NORMDIST	-	0/54	12.8	1E-13	2E-13	31
TDIST	-	0/90	10.1	7E-11	7E-11	64
CHIDIST	-	0/52	13.3	5E-14	3E-14	87
FDIST	-	0/97	13.3	5E-14	3E-13	73
BETADIST	-	0/156	13.4	3E-14	1E-14	74

## 2 Accuracy of the statistical functions

### 2.1 Generalities

As indicated above, the most recent contribution for Excel 2007 is Yalta (2008). Through a series of tables, each devoted to a specific function for computing either probabilities or quantiles for some of the most useful distributions, Yalta (2008) shows that Excel 2007 has generally been improved with respect to previous versions but that the accuracy of its statistical functions is generally weaker than those of OpenOffice.org Calc 2.3 and worse than Gnumeric which gave the correct values in all the cases.

The main problems with Excel 2007 are that, sometimes, either it gives wrong values (at the specified number of significant digits), or non-numeric (e.g. #NUM!) error codes, or, worse still, extreme default values (e.g. 1.0E+7 instead of bigger numbers).

We have built a spreadsheet using the same function calls as Yalta (2008) and examined the results in Excel 2003, 2007 and 2010, as well as in Calc 3.0 and 3.3. Yalta (2008) presented the output of the functions with a certain number of digits, comparatively with the results of ELV Ed. 2 (Knüsel, 2003) which has been used as a standard in several earlier studies. Here, we preferred to show LRE's, see (1), on the basis of Mathematica 7 results. Although ELV and Mathematica results are always very close, the latter, being used under the recommendations of McCullough (2000) (including the use of rational input

**Table 2** Cumulative binomial probabilities at  $k$  for exponent  $n = 1030$  and probability for a success  $p = 0.5$ 

Yalta's Table 2	=BINOMDIST(k;1030;0.5;TRUE)	New name: BINOM.DIST				
$k$	Excel 2003	Excel 2007	Excel 2010	Calc 3.0	Calc 3.3	
1	0.0	0.0	14.9	15.7	15.7	
2	0.0	0.0	13.0	15.7	15.7	
100	0.0	0.0	12.8	15.2	15.2	
200	0.0	0.0	13.4	15.4	15.4	
300	0.0	0.0	14.0	14.7	14.7	
390	0.0	0.0	13.7	14.5	14.5	
391	0.0	0.0	15.1	14.7	14.7	
400	2.1	2.1	13.7	14.8	14.8	
410	4.0	4.0	14.5	14.7	14.7	
420	5.7	5.7	14.5	14.9	14.9	
500	14.0	14.0	15.1	14.8	14.8	
550	14.5	14.5	15.5	15.0	15.0	
575	14.4	14.4	15.1	14.8	14.8	
589	14.4	14.4	15.1	14.8	14.8	

**Table 3** Probabilities of some hypergeometric distribution at point  $k$ , see text for details

Yalta's Table 3	=HYPGEOMDIST(k;500;515;1030)	New name: HYPGEOM.DIST				
$k$	Excel 2003	Excel 2007	Excel 2010	Calc 3.0	Calc 3.3	
0	0.0	0.0	13.2	14.3	14.3	
100	0.0	0.0	13.4	14.5	14.5	
187	0.0	0.0	13.7	14.6	14.6	
188	14.2	14.2	13.3	14.4	14.4	
200	14.2	14.2	13.4	14.2	14.2	
300	14.2	14.2	13.4	14.2	14.2	
312	14.2	14.2	13.3	14.4	14.4	
313	0.0	0.0	13.7	14.6	14.6	
400	0.0	0.0	13.4	14.5	14.5	
500	0.0	0.0	13.2	14.3	14.3	

and `$MaxExtraPrecision = 100`, and we have even used 1000) are given with more digits than the former. Tables 2 to 10 of this paper correspond to Yalta (2008) Tables 2 to 10. When an error occurred in an evaluation, the obtained cell content is put instead of LRE. Cases entries which were added with respect to Yalta (2008) are shaded. This is similar to the supplemental material of Keeling and Pavur (2011) except that we show LRE's instead of function values.

Although it was not said explicitly, Yalta (2008) Tables 2 to 10 are based on a much bigger set of test cases (shown in Table 1 column # cases in file

**Table 4** Simple (if sigma is FALSE) or cumulative (if sigma is TRUE) Poisson probabilities at  $k$  for different values of mean parameter  $\lambda$

Yalta's Table 4		=POISSON(k;lambda;sigma)			New name: POISSON.DIST		
$k$	lambda	sigma	Excel 2003	Excel 2007	Excel 2010	Calc 3.0	Calc 3.3
0	200	FALSE	0.0	0.0	14.3	14.3	14.3
103	200	FALSE	0.0	0.0	14.3	14.7	14.7
104	200	FALSE	13.2	13.2	14.0	14.3	14.3
133	200	FALSE	13.1	13.1	14.4	14.1	14.1
134	200	FALSE	13.2	13.2	14.4	14.6	14.6
200	200	FALSE	13.2	13.2	15.2	14.7	14.7
314	200	FALSE	13.2	13.2	14.0	14.8	14.8
315	200	FALSE	0.0	0.0	13.8	14.3	14.3
400	200	FALSE	0.0	0.0	14.1	14.8	14.8
900	200	FALSE	0.0	0.0	13.9	14.4	14.4
1.E+03	1.E+03	TRUE	9.4	9.4	14.8	#VALUE!	13.2
1.E+05	1.E+05	TRUE	0.0	0.0	15.4	#VALUE!	10.1
1.E+07	1.E+07	TRUE	0.0	0.0	16.0	#VALUE!	#N/A
1.E+09	1.E+09	TRUE	0.0	0.0	15.7	#VALUE!	#N/A
600	200	FALSE	0.0	0.0	13.1	14.7	14.7

**Table 5** Distribution function at  $x$  of a  $\Gamma(\alpha, 1)$  distribution

Yalta's Table 5		=GAMMADIST(x;alpha;1;TRUE)			New name: GAMMA.DIST		
$x$	alpha		Excel 2003	Excel 2007	Excel 2010	Calc 3.0	Calc 3.3
0.1	0.1		#NUM!	#NUM!	15.3	15.0	15.0
0.2	0.1		6.5	6.5	15.1	15.0	15.0
0.2	0.2		6.1	6.1	15.8	15.8	15.8
0.3	0.2		6.5	6.5	15.1	15.4	15.4
0.3	0.3		6.2	6.2	15.5	15.0	15.0
0.4	0.3		6.4	6.4	14.9	14.7	14.7
0.4	0.4		6.3	6.3	15.2	15.8	15.8
0.5	0.4		6.5	6.5	15.1	16.0	16.0
0.5	0.5		6.3	6.3	14.8	14.8	14.8
0.6	0.5		6.5	6.5	14.9	14.9	14.9

before the “+”) than those published (shown in Table 1 column # cases in table). At the request of a referee of a previous version of this paper, we asked and received the original Excel workbook with the results for Excel 2007 but also Calc 3.1 and Gnumeric 1.7.11, among others. Most Mathematica values were not given with enough digits so we had to compute them. We have added some entries (shown in Table 1 column # cases in file *after* the “+”) to cover more extreme cases. We also added values obtained from Excel 2010 and Calc 3.3 and sent the updated file to its original author. A synthesis of the functions

**Table 6** Quantile of order  $p$  of a standard normal distribution

Yalta's Table 6 $p$	=NORMINV( $p;0;1$ )		New name: NORM.INV			
	Excel 2003	Excel 2007	Excel 2010	Calc 3.0	Calc 3.3	
0.49999999999	5.2	5.2	5.2	5.2	5.2	5.2
5.00E-01	15.0	15.0	15.0	15.0	15.0	15.0
1.00E-01	15.5	15.5	15.3	15.2	15.2	15.2
1.00E-02	14.4	14.4	15.4	15.4	15.4	15.4
1.00E-03	14.1	14.1	15.0	15.0	15.0	15.0
1.00E-04	12.4	12.4	16.0	16.0	16.0	16.0
1.00E-05	11.1	11.1	14.9	14.9	14.9	14.9
1.00E-06	8.2	8.2	14.7	14.7	14.7	14.7
1.00E-07	12.9	12.9	14.9	14.9	14.9	14.9
1.00E-15	15.1	15.1	15.2	15.2	15.2	15.2
1.00E-16	15.2	15.2	15.2	15.2	15.2	15.2
1.00E-100	15.0	15.0	15.1	15.1	15.1	15.1
1.00E-197	6.8	6.8	6.8	6.8	6.8	6.8
1.00E-198	2.8	2.8	6.8	6.8	6.8	6.8
1.00E-300	0.0	0.0	7.0	7.0	7.0	7.0

treated and number of cases in Tables 2 to 10 and in the bigger original file is shown in Table 1. We also added four other columns:

- the minimum LRE; note that the cases causing small LRE values were added to the specific tables (Tables 6, 7 and 9);
- the maximum relative difference between Excel 2010 and Mathematica;
- the maximum relative difference between Calc 3.3 and Mathematica;
- the percentage of differences between Excel 2010 and Calc 3.3 equal to 0.

The maximum was taken over all test cases. When the true value was 0, an absolute difference was taken instead of the relative difference. Some cases had to be omitted when one of the programs gave an error. The last four columns of Table 1 will be discussed in the conclusions of this Section.

## 2.2 Binomial distribution (Yalta's Table 2)

For the cumulated binomial probabilities considered, Table 2 shows that Excel is clearly improved in version 2010 compared to the preceding versions of 2003 and 2007. As said above, it is confirmed on a wider set of text cases. Of course this is not a guarantee that all possible probabilities are correct. We will not repeat this warning for all the functions being treated.



**Table 7** Quantile of order  $1 - p$  of a  $\chi^2$  distribution with  $n$  degrees of freedom

Yalta's Table 7		=CHIINV(p;n)		New name: CHISQ.INV.RT		
$p$	$n$	Excel 2003	Excel 2007	Excel 2010	Calc 3.0	Calc 3.3
2.00E-01	1	6.4	6.4	14.4	14.5	14.5
2.00E-01	5	8.1	8.1	15.9	15.9	15.9
1.00E-01	1	6.7	6.7	14.8	15.0	15.0
1.00E-01	5	8.4	8.4	15.1	14.9	14.9
1.00E-05	1	9.6	9.6	12.2	14.4	14.4
1.00E-05	5	10.8	10.8	16.0	15.6	15.6
1.00E-06	1	9.2	9.2	11.2	14.9	14.9
1.00E-06	5	9.8	9.8	14.7	14.7	14.7
1.00E-07	1	#NUM!	#NUM!	10.4	14.8	14.8
1.00E-07	5	#NUM!	#NUM!	15.3	15.2	15.2
1.00E-12	1	#NUM!	#NUM!	6.1	15.0	15.0
1.00E-12	5	#NUM!	#NUM!	15.1	15.4	15.4
4.80E-01	778	#NUM!	#NUM!	14.3	13.5	13.5
5.00E-01	780	7.1	7.1	16.0	15.8	15.8
5.20E-01	782	#NUM!	#NUM!	16.0	14.4	14.4
1.00E-09	1.E+07	#NUM!	#NUM!	16.0	#N/A	#N/A
1.00E-12	1.E+09	#NUM!	#NUM!	14.3	#N/A	#N/A
1.00E-13	1	#NUM!	#NUM!	4.5	15.0	15.0

### 2.3 Hypergeometric distribution (Yalta's Table 3)

Table 3 shows that Excel 2010 is improved with respect to Excel 2003 and 2007 since it gets all the probabilities right, at least for the combinations of the three integer parameters under consideration, Knüsel (1998). At the displayed accuracy, it gets the same values as Calc 3.0 and 3.3. Note that a first attempt failed because the three integers are to be entered in the given order, not in the order given in Yalta (2008) Table 3 legend. Knüsel (2011) has noticed that, for some combination of the parameters (e.g.  $k = 510^8$ ,  $N = 210^9$ ,  $M = 10^9$ ,  $n = 10^9$ ), computing time is surprisingly large, greater than one minute. We were able to reproduce that problem.

### 2.4 Poisson distribution (Yalta's Table 4)

As shown in Table 4, at least for the tested cases, the Poisson probabilities of Excel 2010 are better than those of versions 2003 and 2007, by giving the exact result, up to the depicted accuracy. It does even better than OpenOffice.org Calc 3.0 and 3.3. Version 3.3 of Calc is slightly better than version 3.0 whose errors apparently prevented users from saving the workbook. Knüsel (2011) has mentioned some discrepancies but, as Lewis (2012, personal communication) points out, probabilities are then of order  $10^{-306}$ , close to the smallest entry

**Table 8** Quantile of order  $p$  of a beta distribution over  $[0; 1]$  with parameters 5 and 2

Yalta's Table 8	=BETAINV(p;5;2;0;1)	New name: BETA.INV				
$p$		Excel 2003	Excel 2007	Excel 2010	Calc 3.0	Calc 3.3
1.00E-01		6.9	6.9	14.9	15.1	15.0
1.00E-02		5.2	5.2	14.8	14.8	14.7
1.00E-03		4.2	4.2	15.8	16.0	15.8
1.00E-04		3.2	3.2	14.4	14.4	14.4
1.00E-05		2.2	2.2	15.7	15.7	15.4
1.00E-06		1.5	1.5	14.7	14.7	14.7
1.00E-07		0.0	0.0	14.8	14.8	14.9
1.00E-08		0.0	0.0	14.6	14.5	14.5
1.00E-09		0.0	0.0	14.1	14.0	14.1
1.00E-10		0.0	0.0	14.9	14.8	14.9
1.00E-11		0.0	0.0	14.7	14.6	14.7
1.00E-12		0.0	0.0	14.4	14.4	14.4
1.00E-13		0.0	0.0	14.4	14.3	14.4
1.00E-100		0.0	0.0	14.6	0.0	15.4

in Excel  $2^{-1022} = 2.2251 \cdot 10^{-308}$ . This is the case especially in Kn usel (2011)'s Table 1, where most of the probabilities are smaller than that number. We will not repeat that remark for other cases with such small probabilities.

## 2.5 Gamma distribution (Yalta's Table 5)

For the gamma distribution shown in Table 5, Excel 2010 appears to be better than its predecessors by giving the exact result in all the considered cases. It reaches the accuracy of OpenOffice.org Calc 3.0 and 3.3.

## 2.6 Normal distribution (Yalta's Table 6)

As shown in Table 6, Excel 2010 does better than Excel 2003 and 2007 for most of the quantiles of the normal distribution, by giving the exact result for the extreme cases. It is thus close to OpenOffice.org Calc 3.0 and 3.3. Slightly surprisingly, probabilities very close to 0.5 are the most prone to errors. As is noticed by Lewis (2012, personal communication), the first row is somewhat misleading because the number entered in Excel does not have an exact binary representation and is truncated and in such a case we should not copy and paste numbers from Mathematica to Excel but rather export Mathematica output as an Excel file before computing LRE's. Using numbers of the form  $0.5 - 2^{-j}$  instead and following that procedure, the agreement was better in general but nevertheless we obtained  $LRE = 2.4$  for  $j = 29$ .

**Table 9** Quantile of order  $1 - p/2$  of a Student distribution with  $n$  degrees of freedom

Yalta's Table 9		=TINV(2*p;n)		New name: T.INV.2T		
$p$	$n$	Excel 2003	Excel 2007	Excel 2010	Calc 3.0	Calc 3.3
2.00E-01	1	9.2	9.2	14.6	9.4	14.8
1.00E-01	1	9.7	9.7	14.9	9.4	16.0
1.00E-02	1	9.8	9.8	14.7	14.4	15.5
1.00E-03	1	9.8	9.8	14.8	11.1	15.0
1.00E-04	1	9.8	9.8	15.5	14.8	15.5
1.00E-05	1	9.8	9.8	14.6	0.2	14.6
1.00E-06	1	9.8	9.8	14.8	0.0	14.9
1.00E-07	1	9.8	9.8	15.2	0.0	15.1
1.00E-08	1	0.0	0.0	14.8	0.0	15.0
1.00E-09	1	0.0	0.0	14.7	0.0	14.6
1.00E-10	1	0.0	0.0	14.9	0.0	14.5
1.00E-11	1	0.0	0.0	14.4	0.0	14.3
1.00E-12	1	0.0	0.0	13.1	0.0	13.1
1.00E-13	1	0.0	0.0	12.1	0.0	12.1
1.00E-100	1	0.0	0.0	14.7	0.0	13.8
1.00E-12	1000	5.4	5.4	5.4	5.4	5.4
1.00E-12	10000	5.4	5.4	5.4	5.4	5.4
1.00E-12	1.E+07	10.2	10.2	11.3	#VALUE!	#VALUE!

### 2.7 Chi-squared distribution (Yalta's Table 7)

With regards to the specific quantiles of a chi-squared distribution illustrated in Table 7, Excel 2010 is again much better than its predecessors. It approaches OpenOffice.org Calc 3.0 and 3.3. The weakest results are for very small probabilities and no longer large degrees of freedom. Knüsel (2011) notes some problems with  $p \leq 0.08$  and  $n = 10^{10}$ . As seen in Table 7, in two cases ( $p = 10^{-9}$ ,  $n = 10^7$ , and  $p = 10^{-12}$ ,  $n = 10^9$ ) neither Calc 3.0 nor Calc 3.3 could provide a result whereas Excel 2010 LRE's are respectively 16.0 and 14.3. Moreover these results show inconsistency between CHISQ.INV.RT and CHISQ.DIST. On the contrary for  $p = 10^{-13}$ ,  $n = 1$ , LRE is 4.5 for Excel 2010 but 15.0 for the two versions of Calc.

### 2.8 Beta distribution (Yalta's Table 8)

About the specific quantiles of the beta distribution depicted in Table 8, Excel 2010 is much better than Excel 2003 and 2007. It scores as well as OpenOffice.org Calc 3.3 for which an inaccuracy of version 3.0 was corrected. Knüsel (2011) has noticed some chaotic results for values of  $b$  greater than  $10^{13}$  or  $10^{14}$ .

**Table 10** Quantile of order  $1 - p$  of a Fisher-Snedecor distribution with  $m$  and  $n$  degrees of freedom

Yalta's Table 10	=FINV(p;m;n)			New name: F.INV.RT				
	$p$	$m$	$n$	Excel 2003	Excel 2007	Excel 2010	Calc 3.0	Calc 3.3
5.00E-01	1	1	14.7	14.7	14.7	8.5	14.7	
4.00E-01	1	1	8.9	8.9	14.3	9.1	15.0	
2.00E-01	1	1	9.4	9.4	14.9	9.1	14.4	
1.00E-01	1	1	9.5	9.5	14.6	8.4	14.9	
1.00E-02	1	1	9.5	9.5	14.6	8.0	15.3	
1.00E-03	1	1	9.5	9.5	16.0	12.0	14.5	
1.00E-04	1	1	9.5	9.5	14.6	15.3	15.4	
1.00E-05	1	1	0.0	0.0	14.6	0.0	15.2	
1.00E-06	1	1	0.0	0.0	16.0	0.0	14.9	
1.00E-12	1	1	0.0	0.0	11.9	0.0	11.9	
1.00E-13	1	1	0.0	0.0	11.0	0.0	11.0	
1.00E-100	1	1	0.0	0.0	14.3	0.0	13.5	
1.00E-12	100000	1.E+05	10.8	10.8	14.0	11.8	14.0	
1.00E-12	1E+06	1.E+06	11.3	11.3	14.0	12.4	13.7	
1.00E-12	1.E+07	1.E+07	11.8	11.8	14.0	11.6	13.2	

## 2.9 Student $t$ distribution (Yalta's Table 9)

For the tested quantiles of a Student distribution, Table 9 reveals that Excel 2010 is much better than its predecessors. The accuracy of OpenOffice.org Calc 3.3 is reached, given that some errors of Calc version 3.0 were corrected. All the tested spreadsheet programs have some problems with small probabilities and large number degrees of freedom. Kn usel (2011) also reports wrong results for  $p = 10^{-80}$  and  $459 \leq n \leq 523$ .

## 2.10 $F$ distribution (Yalta's Table 10)

Concerning the quantiles of a Fisher-Snedecor distribution in the tails, Excel 2010 does a little better than its predecessors as shown in Table 10. It attains the precision of OpenOffice.org Calc 3.3 for which some errors of version 3.0 were corrected.

## 2.11 Conclusion for statistical distributions

We have observed few systematic anomalies in test cases other than those covered by Yalta (2008) tables (which would have occurred if Microsoft had only corrected the cases showing a failure in that paper). On the basis of these results, Microsoft Excel 2010 appears as good as OpenOffice.org Calc 3.3. Calc 3.3 fails on two Poisson probabilities, on some  $\chi^2$  and Student quantiles that

**Table 11** New functions for statistical distributions in Excel 2010. Y means “Yes” and N means “No”, while “N<sup>b</sup>” means that it is not good and “N<sup>f</sup>” that it is feasible by symmetry.

Distribution	prob./ density	Tail probabilities			Quantiles		
		left	right	two-tail	left	right	two-tail
Binomial	Y	Y	N <sup>b</sup>	N	Y	N	N
Hypergeometric	Y	Y	N <sup>f</sup>	N	N	N	N
Poisson	Y	Y	N <sup>b</sup>	N	N	N	N
Gamma	Y	Y	N <sup>b</sup>	N	Y	N	N
Normal	Y	Y	N <sup>f</sup>	N	Y	N <sup>f</sup>	N
$\chi^2$	Y	Y	Y	N	Y	Y	N
Beta	Y	Y	N <sup>f</sup>	N	Y	N	N
<i>t</i>	Y	Y	Y	Y	Y	N <sup>f</sup>	Y
<i>F</i>	Y	Y	Y	N	Y	Y	N <sup>f</sup>

Excel 2010 could compute whereas Excel is not accurate enough for some  $\chi^2$  quantiles. By the way, fractional degrees of freedom are still truncated to integers. That implies that users interested in fractional degrees of freedom should probably look elsewhere.

Note the addition of new functions in Excel 2010 with more explicit names. For more details, see Microsoft (2009a). Obviously that can raise problems of incompatibility when an Excel 2010 workbook is opened in an old version of the software (or in OpenOffice.org Calc, for that matter) since the new names will not be recognized. For reasons of compatibility, Microsoft has hopefully maintained the old names (which we employed here for compatibility with the other programs, mentioning also corresponding new names) while improving accuracy. Please note that CHISQ.INV, T.INV and F.INV now refer to quantiles for given probabilities, contrarily to CHIINV (CHISQ.INV.RT), TINV (T.INV.2T), and FINV (F.INV.RT).

To conclude, most of the problems of Excel raised by Yalta (2008) were corrected in the 2010 version. Some of the remaining problems may be due to the limitations of IEEE-754 arithmetic, although some algorithms have been used to improve the calculations, like TWOSUM to compute the rounding error when adding two numbers, see Microsoft (2009a).

These results have been confirmed by a verification in the more detailed file from which Yalta (2008) is just a sample. Indeed, Table 1 shows, for each statistical function in Yalta file, the maximum of the absolute values of the relative difference between either Excel 2010 or Calc 3.3, on the one hand, and Mathematica 7, on the other. It can be seen that the order of magnitude is smaller than 2.E-5, meaning that 5 correct significant digits are obtained. As mentioned above, that does not mean that all values are correct. First, just the main statistical functions have been considered and the proliferation of new functions may cast a doubt. Second, not all combinations of arguments have been tried since their number is practically unlimited. Third, we have seen that big values of the degrees of freedom of the  $\chi^2$  quantiles are not taken into account correctly (although limiting distribution arguments can be used

in such a case) and this indicates insufficient testing. Fourth, past track of Microsoft failures in the domain of statistical functions is not in their favor.

We have added some additional cases not covered by Yalta (2008), mainly with very small probabilities and high degrees of freedom for  $t$ ,  $\chi^2$  and  $F$  distributions without too many problems. By the way, a close examination of Yalta file completed for Excel 2010 and Calc 3.3 reveals that the differences between results of Excel 2010 and Calc 3.3 are very often exactly zero, except for the hypergeometric probabilities. This is clearly indicated in the last column of Table 1 where the percentage of exact zero differences is given. In overall functions and test cases, the smallest non zero difference is  $-7.29 \cdot 10^{-301}$ . Since it is known that even the two associative laws and the distributive law,  $a + (b + c) = (a + b) + c$ ,  $a * (b * c) = (a * b) * c$  and  $a * (b + c) = (a * b) + (a * c)$  are not practically verified in floating-point arithmetic (see e.g. Ueberhuber, 1997, p. 143). That means that at least some parts of the programs are really very close.

In principle (Microsoft, 2009a) the exponential, lognormal negative binomial and Weibull distributions are also improved. Otherwise, there is no guarantee that the statistical functions other than those considered in Table 1 will be correct. For example, we have checked that the function LOGEST studied by McCullough and Heiser (2008) has not been changed. MULTINOMIAL is not considered as a statistical function but would need improvement. The main criticism raised by Kn usel (2011) and Lewis (2012, personal communication) is that some functions are missing. As shown in Table 11, only the  $\chi^2$  and  $F$  distributions have functions for both left and right tail probabilities and for both left and right quantiles. It is known that  $P(X > x) = 1 - P(X \leq x)$  in theory but in practice, if  $P(X \leq x) \geq 1 - 10^{-14}$ , then  $P(X > x)$  loses accuracy. Microsoft seems to have taken care of this in some cases but not all. Also, among the discrete distributions, the (left) quantile is available only for the binomial distribution. Kn usel (2011) has also noticed inconsistencies between some functions of Excel for tail probabilities and for quantiles, especially for the gamma and the right tail  $\chi^2$  probabilities. Let us now look if improvements also occur in the other areas.

### 3 Generator of pseudo-random numbers

#### 3.1 Generalities

As noted by McCullough and Wilson (2005) there are actually two generators of pseudo-random numbers in the various versions of Excel, and even three if we take Visual BASIC for Application (VBA) into account:

- (A) the generator called upon by function RAND() and used when generating other pseudo-random deviates with other distributions than the uniform law over  $[0; 1[$ ;
- (B) the generator provided by the Statistical Toolpak add-in, e.g. by going to Tools > Data Analysis > Random Number Generation in Excel 2003;

- (C) the generator used in VBA `rnd` function.

In Excel 2007, the generator (A) is, in theory, the one of Wichmann and Hill (1982). But McCullough (2008b) has proved that it is not Wichmann and Hill (1982) that has been implemented by Microsoft since Excel 2003. It is also recalled that the period of the generator is  $6.95 \cdot 10^{12}$ , not  $10^{12}$ , as reported in the original article. McCullough and Wilson (2005) showed that (B), which has not been changed in Excel 2003, can easily give 0 (a little surprising) or 1 (very astonishing) which indicates an implementation error in the algorithm. The algorithm (C) is known to be bad, see L'Ecuyer and Simard (2007), with a period of only  $2^{24} = 16.10^6$  and bad properties.

Several authors, e.g. McCullough and Wilson (2005), McCullough (2008b), have requested a generator with a longer period and with better properties in spreadsheet processors. A well-recognized way to test generators is to use the batteries of tests provided in TestU01, L'Ecuyer and Simard (2007).

### 3.2 Excel 2010

As indicated by Keeling and Pavur (2011), Microsoft (Oppenheimer, 2009a) states to have improved the RAND function in Excel 2010. Before investigating that claim, let us look at generators (B) and (C). The left part of Fig. 1, obtained by following the instructions of McCullough and Wilson (2005), with the Data Analysis Toolpak Random Number Generation tool, shows

1. that the results obtained are identical to those of Excel 2007 and before;
2. that the value “1” is still there.

We conclude that the generator (B) of the Statistical Toolpak add-in is unchanged, and therefore has the same failures as in the previous versions. For the generator (C) used in VBA, in the same way, we empirically checked (by using the `rnd` function to fill a column of a worksheet and then copying the obtained values), that the results are identical to those of Excel 2003. See the right hand side part of Fig. 1. Hence, there is no improvement in this area either.

Let us now check the new generator (A). We applied a procedure similar to that proposed by McCullough (2008b), sections 3 and 4, who performed high-precision calculations with the Python language. Here, we carried out our calculations with a FORTRAN program making use of a multiprecision library (Bailey, 1993). We could reproduce the checking procedure carried out by McCullough (2008b) on the generator of Wichmann and Hill (1982), as it is implemented in R, being able to recover the subsequent pseudo-random number deviates. On the contrary, the procedure failed on a sample of random numbers produced by the generator of Excel 2010. The explanation, which is currently semi-official, see the blog Oppenheimer (2009a), is that the new generator makes use of the Mersenne Twister algorithm, Matsumoto and Nishimura (1998). Such a change is a request of several authors, e.g. McCullough and Wilson (2005), McCullough (2008b). That generator passes most

**Fig. 1** Excerpts of sequences of pseudo-random numbers obtained in Excel 2010 by the add-in Random Number Generator and by the VBA function `rnd` in Excel 2003 and 2010

	A	B
1	Excel 2003	Excel 2010
2	0,7055475	0,7055475
3	0,533424	0,533424
4	0,5795186	0,5795186
5	0,2895625	0,2895625
6	0,301948	0,301948
7	0,7747401	0,7747401
8	0,0140176	0,0140176
9	0,7607236	0,7607236
10	0,81449	0,81449
11	0,7090379	0,7090379
12	0,0453528	0,0453528
13	0,4140327	0,4140327
14	0,8626193	0,8626193
15	0,79048	0,79048
16	0,3735362	0,3735362

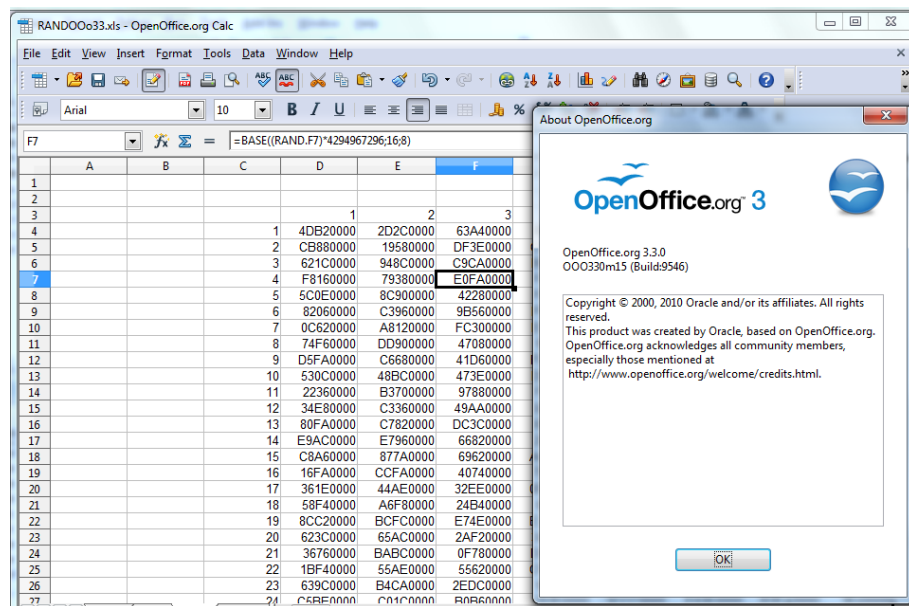
tests in L'Ecuyer and Simard (2007) and its period is  $2^{19937} - 1$ . L'Ecuyer and Simard (2007) mention that it passes all SmallCrush tests and all but two tests of the bigger batteries of empirical tests called Crush and BigCrush. Simard (2011, personal communication) has mentioned that the two tests failed with version 1.0 are Linear Complexity,  $r = 1$ , and Linear Complexity,  $r = 29$ , with a p-value  $< 10^{-15}$ . McCullough (2006) has tested all the random number generators in version 1.6.1 of the software package 'R' and has concluded that the Mersenne Twister generator passes all the tests of the three batteries of tests in version 1.1.

Taking into account the previous implementation problems, it should be wise to examine the quality of the new generator (A). To test the quality of the new random number generator in Excel 2010, the only way is to generate a flow of pseudo random numbers and enter it in a software for testing random number generators, like TestU01, L'Ecuyer and Simard (2007). Unfortunately, that takes time. The Crush battery requires about  $2^{35}$  numbers, i.e. a file of about 412GB and the BigCrush test about eight times more. In order to test the Excel 2010 in a reasonable span of time, we generated a  $32 \times 1024$  range at a time and modified the Crush battery so that the file is rewound after each test.

In practice, we created a new CrushFile battery on the basis of the SmallCrushFile battery provided in the implementation of TestU01 version 1.2.3 and the Crush battery. Generation of the test file took about two weeks and



**Fig. 2** Result of hexadecimal conversion of a sequence of pseudo-random numbers obtained in Calc 3.3 using RAND function



the test itself required 36 hours of computation (on a computer with a Pentium 4 processor clocked at 2GHz). All tests are passed except Periods in Strings with  $r = 15$  and  $s = 15$  for which the  $p$ -value is  $8.10^{-7}$ . The difference with L'Ecuyer and Simard (2007) results can be explained by the simultaneous generation of a whole range at a time and the use of the same stream for each test to restrict the file to a manageable size. Given the information in L'Ecuyer and Simard (2009), the failed test, Guibas and Odlyzko (1981), omits the first  $r$  ( $= 15$ ) significant bits of the number and uses the next  $s$  ( $= 15$ ) bits for the test. So if the user doesn't put too much emphasis on the lower bits, we don't think the problem is serious. To be more specific, imagine that we generate two variables  $X$  and  $Y$  from a single pseudo-random number  $U$ , taking  $X = U$  and  $Y = (2^{15}) * U$ . Then the generated values for  $X$  will be acceptable but those of  $Y$  will show some undesirable patterns.

Finally let us note that, to avoid the very bad VBA `rnd` function, it should be nice to use Excel functions in VBA macros, therefore in particular the possibly good new generator (A) instead of the surely defective generator (C). This is not possible by using `WorksheetFunction.RAND()` but, as Lewis (2012, personal communication) indicates, by using `Evaluate("RAND()")`.

### 3.3 Calc 3.3

To our knowledge, aside from a paragraph in Almiron et al. (2010), there has been no study of the pseudo-random number generator used in OpenOffice.org Calc which is moreover not well documented. However, some messages in forums seem very critical. We have reproduced the experiment of Crabtree (2009) on a computer equipped with Windows XP and OpenOffice.org Calc 3.3, which was performed on an unspecified version of Calc. The observations reproduced in Fig. 2 are identical. Indeed, after multiplication by  $2^{32}$ , the numbers seen in hexadecimal representation all end up with “0000”, showing only 32-bit accuracy. According to messages in forums, it is known that the maximum number we can reach of 0.9999695. A finer analysis reveals that, as a matter of fact, there are 32,768 different values, so only a 15-bit accuracy.

It is not easy to know more, as Almiron et al. (2010) have already noticed. They indicate that the random number generators in Calc depend on the standard C library implemented in the development tools, hence is non portable a priori. According to OpenOffice.org instructions to developers, Microsoft Visual C++ 2008 Express should be used for the Windows version but it is no longer available. The limitation to 15-bit accuracy is confirmed by a check of the value 32,768 of `RAND_MAX` in the `stdlib.h` file of Microsoft Visual C++ 2010 Express. On the contrary, our investigations with users of recent Linux versions of OpenOffice.org show that they benefit from 2,147,483,647 different values instead of 32,768 for users working under Windows. Hence the maximum value we can reach of 0.999999995343.

We made some experiments (generation of random numbers in a  $2048 \times 128$  array) that indeed confirm these suspected differences. This could be verified in a Mandriva 2009 Linux installation using Calc 3.1 and on an Ubuntu 10.10 installation using Calc 3.2 (where we got 0.99999694061 as maximum, thus greater than 0.9999695). Of course, the corresponding analogue of Fig. 2 shows hexadecimal numbers that do not end with “0000”.

The advantage of a free source software is that, in principle, anybody could do it for his/her own usage and make it profitable to other people. During our investigations, an extension MTRNG of Calc appeared from contributor Hanya, with a Mersenne Twister algorithm, Matsumoto and Nishimura (1998).

### 3.4 Gnumeric 1.10.16

Gnumeric has made use of the same Mersenne Twister algorithm for several years, apparently since the McCullough (2004c, pg. 8-9) remarks.

### 3.5 Conclusions for pseudo-random numbers

The new random number generator `RAND()` in Excel 2010 is much more satisfactory than in the previous versions and is similar to Gnumeric in this

respect. The VBA rnd generator should be replaced by RAND as suggested by Lewis (2012), see above. The situation is more mitigated for OpenOffice Calc 3.3 since the quality of the pseudo-random generator depends on the operating system. It is bad for the Windows platform. Note that a Mersenne Twister algorithm is now included in OpenOffice.org 4.0 and in LibreOffice 4.0.

## 4 Quality of the statistical add-ins

### 4.1 Generalities

The statistical add-ins were available in Tools > Data Analysis in Excel 2003 and are now in the Data ribbon. McCullough and Heiser (2008) and Almiron et al. (2010) have analyzed Excel 2007 in depth. The usual tests for that purpose are Wilkinson's tests and StRD datasets.

For further details and background on Wilkinson's tests, see Sawitzki (1994) and McCullough (2004b). These tests work on an artificial data set of 10 observations simulating number patterns that can cause problems in plots and computations. Variable X is such that  $X_i = i$ ,  $i = 1, \dots, 10$ , variables ZERO and MISS equal the constant 0 and a missing value (#N/A in Excel), respectively. The variables BIG, LITTLE, HUGE, TINY and ROUND are defined by  $BIG_i = 99999990 + i$ ,  $LITTLE_i = 0.99999990 + i10^{-8}$ ,  $HUGE_i = i10^{12}$ ,  $TINY_i = i10^{-12}$ ,  $ROUND_i = i - 0.5$ .

Statistical Reference Datasets (StRD) have been developed by the (American) National Institutes of Standards and Technology, see their website <http://www.nist.gov/itl/div898/strd>. These datasets have been examined by McCullough and Heiser (2008) and Almiron et al. (2010) for Excel 2007. To summarize, low difficulty tests are passed but some high difficulty tests (NumAcc2, NumAcc3, NumAcc4 for the standard deviation and lag 1 autocorrelation, SmLs07, SmLs08, SmLs09 for ANOVA models) are failed.

There are other statistical add-ins for Microsoft Excel, the best known being sold by Addinsoft (XLStat), Analyse-it Software (Analyse-it), and R. Fitch Software (WinSTAT). Other statistical solutions with Excel include Statconn (RExcel), Numerical Algorithms Group (NAG Library from Excel), and SAS Institute (JMP).

### 4.2 Excel 2010

There is not much to say about this subject for Excel 2010. Indeed, besides the "Random Number Generation" tool mentioned in Section 3, in concordance with Microsoft objectives (Microsoft, 2009a, first paragraph of second section on p. 6), everything indicates that nothing has changed with respect to Excel 2007, and thus Excel 2003. For example:

- the missing data bug in the "t-Test: Paired Two Sample for Means" procedure for testing equality of means in the case of paired observations, is still there. It was first observed by Simon (2000).

**Fig. 3** Regression by Excel 2010 on Wilkinson IV-C test

Task IV.C					
SUMMARY OUTPUT					
<i>Regression Statistics</i>					
Multiple R		1			
R Square		1			
Adjusted R Square	0.857142857				
Standard Error	6.1672E-16				
Observations	9				
ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	2	60	30	1.58E+32	6.87763E-96
Residual	7	2.66241E-30	3.8E-31		
Total	9	60			
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>
Intercept	-99999990	7.96182E-09	-1.3E+16	5.4E-111	-99999990
BIG	1	7.96182E-17	1.26E+16	5.4E-111	1
LITTLE	0	0	65535	#NUM!	0

- Regression does not accept more than 16 columns of data in the X area although the statistical function LINEST, on which it said to be based, allows more columns.
- Wilkinson’s Test IV-C, which Microsoft Excel 4.0 failed, as described in Sawitzki (1994), meaning that a regression of X on some two linearly dependent variables is accepted.

This last point is worth discussing. Wilkinson Test IV-C involves the regression of X on a constant, BIG and LITTLE. Like in previous versions, Excel 2010 shows results which look strange, see Fig. 3. They are nevertheless compatible with the documentation which says: “LINEST checks for collinearity and removes any redundant X columns from the regression model when it identifies them. Removed X columns can be recognized in LINEST output as having 0 coefficients as well as 0 se’s. If one or more columns are removed as redundant, then df is affected because df depends on the number of X columns actually used for predictive purposes.” That means that Excel has noticed collinearity, has removed variable LITTLE from the model and expressed X as being  $-99999990 + \text{BIG}$ . Indeed a zero is printed for the coefficient of the omitted variable as well as for the corresponding standard error. Note also that the model degree of freedom is 2, as expected. SAS GLM procedure does the same but displays a dot instead of 0.

Anyway, since the other statistical add-ins have apparently not been changed, we have not checked all the StRD datasets. This was done by Keeling and Pavur (2011).

### 4.3 Calc 3.3

Note that OpenOffice.org Calc has no standard equivalent add-in for data analysis although several extensions do exist.

### 4.4 Gnumeric

Gnumeric has a certain number of entries in the Statistics menu which correspond roughly to the offerings in the Data Analysis of Excel plus some other procedures like a few nonparametric tests, a few tests of normality, principal component analysis, Kaplan-Meier analysis and the chi square tests for independence and homogeneity. Those procedures (linear regression and ANOVA) that were tested by McCullough (2004c) succeeded. Testing the other procedures is out of the scope of this paper.

### 4.5 Conclusion for statistical add-ins

The conclusion is that Microsoft did not make an attempt to fix all the errors in Excel, and this point needs to be made strongly. Microsoft continues to market a product that contains known statistical errors, some of them going back to Excel 4, released in 1994. This should be contrasted with what is found on some open source or free software packages, see McCullough (2004c) about Gnumeric and Yalta and Yalta (2010) about gretl.

## 5 Solver nonlinear regression

### 5.1 Generalities

There are more and more uses of optimization algorithms like quasi maximum likelihood estimation, including time series modeling and logistic regression, M-estimation, etc. but they are generally performed using specialized statistical software. But the first use for an optimization algorithm in statistics may be curve fitting. This may be the reason why the National Institutes of Standards and Technology, in their StRD datasets, have developed a suite of 27 nonlinear regression problems of various levels of difficulty: low (l), average (a), high (h), see Table 12 for datasets names and levels. These datasets are available on their website, see above. They all have one explanatory variable except one (Nelson) which has two explanatory variables. The number of parameters varies between 2 and 9 and the number of observations varies

between 6 (BoxBOD and DanWood) and 250 (Gauss1, 2 and 3). For each of these problems there are certified solutions. Each problem has two sets of initial values, denoted Start I (relatively far from the certified solution) and Start II (closer to the certified solution).

The analysis is based on previous studies on general nonlinear regression and more specifically on the StRD datasets. McCullough (1998) has introduced the use of the StRD benchmark tests for statistical packages, measuring the number of correct significant digits of an estimate  $q$  with respect to a correct value  $c$  by LRE as defined in (1). For several parameters, the minimum of the LRE value is determined as the problem LRE.

Fylstra et al. (1998) have described the design and use of the Excel 97 Solver. Note that the Excel Solver is aimed at solving moderately sized (200 variables at most) linear and nonlinear optimization problems, under constraints or not, also with integer programming capability. We will just discuss unconstrained nonlinear problems for which StRD datasets are designed since the other algorithms (the simplex algorithm for linear problems and branch-and-bound algorithm for integer programming) of the Excel Solver are better judged for problems in operations research, although they are used in some specific statistical methods like quantile regression (Koenker, 2005).

McCullough (1999) has illustrated the StRD tests on several statistical packages. McCullough and Wilson (1999) have delivered a similar study on Excel 97. McCullough and Wilson (2002) have updated their study for Excel 2000 and 2002 but have observed no change. Overall, the results are not very satisfactory because 14 out of the 27 problems lead to  $LRE = 0$ . However, Berger (2007) has noticed a programming error in the Excel implementation of three problems (Gauss1, 2 and 3), due to nonstandard operator precedence in Excel, so that 11 problems are still solved with zero digit accuracy. Almiron et al. (2010) have noticed that, when comparing two problems, a graphically better fit can sometimes be obtained with a smaller LRE but we will nevertheless use the problem LRE as the criterion.

For a general discussion of nonlinear estimation, together with general recommendations, see McCullough (2004a), which also provides an illustration on one of the StRD problems (Misra1a). McCullough and Renfro (2000) provide a more detailed account of algorithms while possible causes of failure are discussed. The method used by Excel Solver for nonlinear problems in previous versions was called Generalized Reduced Gradient 2 (GRG2), see Lasdon et al. (1978) for details, with a choice of a Newton (default) and a Conjugate search in a direction, and a provision of automatic scaling. As a consequence of integration in a spreadsheet, no analytic derivatives capability is provided, see Fylstra et al. (1998). Also, standard errors of the estimates are not available.

McCullough and Wilson (1999) have concluded for Excel 97 that the default convergence setting of  $10^{-4}$  was better replaced by  $10^{-7}$  for the StRD problems and recommended the automatic scaling option. For that reason, we will consider these two changes to the default options. For Excel 2007, Almiron et al. (2010) have more or less confirmed the previous studies on the earlier versions although they have used both default options (mimicking the behav-

ior of a user who has no prior idea about the problem, its solution and the algorithm) and customized settings, but different from those recommended by McCullough and Wilson (1999). For example, automatic scaling does not seem to have been used by Almiron et al. (2010). Moreover, they have also added the Conjugate search procedure and present results for Start I and Start II initial values.

On the other hand, McCullough and Wilson (1999) and McCullough and Wilson (2002) have implemented a strict strategy to use Start I (far) initial values but use Start II (close) initial values if the Solver clearly states that it cannot find a solution from Start I initial values. In practice, since the Solver provides results even when an optimum is not found, results with Start II initial values are not shown. At this stage, we have preferred to follow Almiron et al. (2010) and provide a table when Start II initial values are used.

## 5.2 Excel 2010 Solver for nonlinear regression

The main purpose of this Section is to examine the new Solver in Excel 2010. We performed the test on a PC with an Intel i386 processor and a Microsoft Vista 32 bit operating system. That implies that we have not covered the 64 bit version of Office 2010 but this is not really a restriction (Fylstra and Gupta, 2011, personal communication).

Like in previous versions, besides a maximum time limit and a maximum number of iterations, the Solver will stop and display a message “Solver found a solution” (“fnd” in Tables 11 and 12) if GRG2 has found a locally optimal solution, which, for an unconstrained problem and according to the help system, means that “There is no other set of values for the decision variables close to the current values that yields a better value for the objective”. This is unfortunately not quantified. It appears (Fylstra and Gupta, 2011, personal communication) that this message occurs if the  $L_\infty$  norm of the objective gradient is less than or equal to the Constraint Precision tolerance (by default  $10^{-6}$ ). This corresponds to the test for the Kuhn-Tucker conditions. But the Solver will also stop and issue another message “Solver converged to the current solution” (“curr” in Tables 11 and 12) if the absolute value of the relative change in the objective function is less than the Convergence tolerance in the Solver Options dialog (by default  $10^{-4}$ ) for the last 5 iterations, and the test for the Kuhn-Tucker conditions has not yet been satisfied. There is therefore no guarantee of convergence for the different parameters of the nonlinear regression model. That “solution” is nevertheless displayed in the spreadsheet if the user accepts it. The role of the Convergence and Precision settings has not been stated clearly in the past and the lack of details in the help system has not helped. There are some other messages described below.

The interface is revamped, with more detailed and explicit messages. For example, with the second best message “Solver has converged to the current solution”, it is now clearly indicated that “Solver has performed 5 iterations for which the objective did not move significantly” with the recom-

mendation “Try a smaller convergence setting or a different starting point”. Also, there are more details in the reports about the options, see Fig. 4. Excel 2010 help consists of several topics including “Define and solve a problem by using Solver”. It ends up with a link “More help on using Solver” leading online to “Solver Help at [www.solver.com](http://www.solver.com)”. A direct link which is “<http://www.solver.com/excel2010/solverhelp.htm>” is also available outside of Excel. It contains a complete documentation including the new non-deterministic methods and details on all the options. In particular it contains recommendations for the different methods and explanations for the messages issued by the Solver. Like in previous versions, the user is requested to contact Frontline Systems, Inc. for more information on the internal solution process, see Oppenheimer (2009b). Indeed Frontline Systems has made the Solver for Microsoft and sells Premium Solver Pro as an upgrade of the standard Excel Solver. The present author has asked questions to Frontline Systems support and has posted messages in Microsoft forums from which it appears that Frontline Systems supports not only the products they sell but also the standard Excel Solver.

In Excel 2010 Solver, there is a new method in addition to GRG2 called Evolutionary method, and, partly hidden in the options, a Multistart option added to the default GRG2 method. By default both the Multistart option of GRG2 (simply called the Multistart method, from now on) and the Evolutionary method require bounds for the variables (our parameters) but it is possible to avoid specifying these bounds. For the description of these new methods, we follow the documentation and (Fylstra and Gupta, 2011, personal communication).

Fig. 4 Answer report for StRD BoxBOD dataset, Start I initial values, Multistart method

**Microsoft Excel 14.0 Answer Report**  
**Worksheet: [BOXBOD\_s10.xls]BOXBOD I**  
**Report Created: 17/05/2011 10:45:08**  
**Result: Solver converged in probability to a global solution.**

**Solver Engine**  
 Engine: GRG Nonlinear  
 Solution Time: 0.406 Seconds.  
 Iterations: 0 Subproblems: 3

**Solver Options**  
 Max Time 100 sec, Iterations 100, Precision 0.000001, Use Automatic Scaling  
 Convergence 0.000001, Population Size 100, Random Seed 0, Derivatives Forward, Multistart  
 Max Subproblems Unlimited, Max Integer Sols Unlimited, Integer Tolerance 5%, Solve Without Integer Constraints

Objective Cell (Min)

Cell	Name	Original Value	Final Value
\$G\$2	B2 RSS	186382.3817	9771.499976

Variable Cells

Cell	Name	Original Value	Final Value	Integer
\$C\$3	B1	1	172.5000001	Contin
\$D\$3	B2	1	20.62246104	Contin



The Multistart method is designed for smooth nonlinear, non-convex problems. It operates by starting the GRG2 algorithm from a number of different starting points for the variables. These initial values are chosen randomly, within the bounds on the variables. Corresponding local optima are then clustered using a method known as multi-level single linkage, see Rinnooy-Kan and Timmer (1987a) and Rinnooy-Kan and Timmer (1987b). The local searches continue until a Bayesian test that all of the locally optimal solutions have probably been found is satisfied. The method is inherently probabilistic because the method relies on sampling within the hypercube defined by the bounds on the variables.

The Evolutionary Solver is designed for non-smooth, non-convex problems, e.g. when the objective or constraints rely on non-smooth or discontinuous Excel functions such as IF, CHOOSE, or LOOKUP. It is also inherently probabilistic, and it has fewer guarantees on its performance, and more possible error conditions, than Multistart or the GRG2 method alone. It too relies on random sampling in the hypercube defined by the bounds on the variables. Both the Multistart and the Evolutionary method take much more time than the simple GRG2 method. Indeed they are not intended to find a local optimum for GRG2 but a global optimum. To possibly achieve this goal, lots of subproblems are considered with various initial values. Multistart still makes use of GRG2 starting values but the Evolutionary method apparently does not use them at all. These two Multistart and Evolutionary methods can also issue different messages, the best being that “Solver has found a global optimum”, denoted “glob” in Tables in Table 12 and 13.

It is time to introduce our experiments. We have first adopted the two options taken by McCullough and Wilson (1999) and McCullough and Wilson (2002): (i) use Automatic scaling, (ii) use Convergence tolerance of  $10^{-7}$ , instead of the default  $10^{-4}$ . This will be called set 1 of method parameters in Tables 12 and 13, which refer, respectively, to Start I (far) and to Start II (close) initial values. We have used Newton search default only, since the conjugate search option was systematically inferior in the results of Almiron et al. (2010). Columns 2 (GRG2), 4 (Multistart) and 6 (Evolutionary) provide problem LRE’s for these three methods, respectively. Columns 3 and 5 contain either a shortened form of GRG2 Solver message (in addition to “curr” already described), “fnd” meaning that a local optimum was *found*, “glob” meaning that a *global* optimum was found, or “err” if the Solver has issued an *error* message preventing the production of an Answer report, or “big” if a solution is given but showing at least one huge number, typically of order  $10^{30}$ . Column 7 contains the result of the Evolutionary method which is either “init” if optimal values found are simply the *initial* values (with a message that Solver could not improve them), “big” if the solution shows at least a huge number, “noimp” if the algorithm *cannot improve* the solution which is neither the starting point nor a point containing a big coordinate. Columns 8 and 9 refer to Calc Solver for Nonlinear Regression in Section 5.3. Column 10 contains LRE’s for Gnumeric. Columns 11 and 12 refer to set 2 of method parameters suggested by (Fylstra and Gupta, 2011, personal communication)

like set 1 but (iii) with Precision tolerance of  $10^{-7}$ , instead of the default  $10^{-6}$ , and (iv) Central derivatives, instead of the simpler forward derivatives.

Let us start with Start I (far) initial values whose results are shown in Table 12. The observations of previous authors, including Berger (2007), are confirmed for the standard GRG2 algorithm with zero digit accuracy in 12 cases out of 27. The detailed results are not identical to those of McCullough and Wilson (1999) and McCullough and Wilson (2002). Not taking Gauss1, 2 and 3 problems into account, the LRE’s are generally smaller except for Misra1c and Misra1d. The results are also not identical to those of Almiron et al. (2010) who didn’t make use of automatic scaling and employed a slightly smaller Convergence setting of  $10^{-8}$  instead of  $10^{-7}$ .

We were suggested to force Solver to iterate solving until the message “Solver has found a solution” is issued. This is relatively easy to do since, when it stops with the message “Solver has converged to a solution”, it is possible to accept that solution and to return to the Solver dialog box and solve again from these new starting values, and so on. We did that with the Start I initial values but there has been an improvement in only two cases (out of 19), Misra1b (with LRE= 7.2) and Kirby2 (with LRE= 2.7). Clearly, set 2 options leads to an improvement for most data sets with  $LRE \geq 6$  in 16 cases, with “fnd” status well justified and 0 in the remaining cases, this time with various issues (4 times “cur”, 6 misleading “fnd” and only one declared lack of convergence).

We can now discuss the last two Multistart and Evolutionary methods, although we have learned that these methods make weak assumptions about the problem functions and are general, too general, so that they will not work well for the smooth, and even flat, objective functions in the StRD problems. Multistart takes time (often 30 s on a 2 GHz Pentium 4 computer) but very frequently (13 cases out of 27) leads to an error message (telling that an error occurred for some values), thus without improvement and detailed “Answer report” output, while showing the same solution as the GRG2 method, . But also in all but one of the remaining cases (13 cases out of 27) a very optimistic message that a global solution was found while the results were exactly the same as those from GRG2 and often (5 cases out of 13) with 0 digit accuracy. For dataset Eckerle4, a global solution is said to be found but it differs from the GRG2 solution and includes at least one huge number. Multistart is thus not to be recommended for nonlinear regression, in this release at least. Note that these results may not be reproducible because the method makes use of random numbers. It is possible to set a seed to the random number generator but we refrained from doing so. In a few cases, a message appeared saying that the maximum computing time was used and we then asked to continue.

The conclusion is still worse for the Evolutionary method: in 6 out of 27 problems, the initial values are obtained and, in the remaining cases, at least one huge number is obtained in the proposed solution. Given that the Evolutionary method takes even more time (with a median of 50 s) than the Multistart method and could never find a solution, we can safely recommend *not* using it.

**Table 12** Results for three methods of Excel 2010 Solver (GRG2 with two sets of parameters, Multistart, Evolutionary), the two methods of Calc 3.3 (DPES, SCO) and Gnumeric NLSolve from Start I initial values. Abbreviations: “curr” means current, “fnd” means found, “glob” means global, “err” means error, “init” means initial, “noimp” means can’t improve

Test (level)	GRG2 set 1		Multistart		Evolutionary		Calc 3.3		Gnumeric		GRG2 set 2	
	LRE	Note	LRE	Note	LRE	Note	DPES	SCO	NLSolve	LRE	Note	LRE
Misra1a (l)	3.5	curr	3.5	err	0	big	0	5.2	8.7	8.7	8.7	fnd
Chwirut2 (l)	4.3	curr	4.3	err	0	init	5.4	5.6	7.5	8.1	8.1	fnd
Chwirut1 (l)	4.4	curr	4.4	err	0	init	5.7	6.6	8.1	8.1	8.1	fnd
Lanczos3 (l)	0	curr	0	err	0	big	0	0	0	0	0	curr
Gauss1 (l)	4.9	curr	4.9	err	0	init	6.6	6.8	8.6	9.2	9.2	fnd
Gauss2 (l)	4.6	curr	4.6	err	0	init	6.2	0	8.4	9.8	9.8	fnd
DanWood (l)	5.1	curr	5.1	glob	0	big	4.6	5.5	8.7	9.4	9.4	fnd
Misra1b (l)	3.5	curr	3.5	glob	0	big	0	0	7.8	8.3	8.3	fnd
Kirby2 (a)	1.1	curr	1.1	glob	0	big	5.1	0	5.4	6.6	6.6	fnd
Hahn1 (a)	0	curr	0	glob	0	big	3.8	0	0	0	0	*
Nelson (a)	0	fnd	0	err	0	noimp	0	0	5.1	7.4	7.4	curr
MGH17 (a)	0	fnd	0	err	0	big	2.0	0	0	0	0	fnd
Lanczos1 (a)	0	curr	0	err	0	big	0	0	0	0	0	curr
Lanczos2 (a)	0	curr	0	err	0	big	0	0	0	0	0	curr
Gauss3 (a)	4.0	curr	4.0	err	0	init	6.1	0	8.1	9.3	9.3	fnd
Misra1c (a)	6.3	curr	6.3	glob	0	big	5.6	5.8	8.2	7.8	7.8	fnd
Misra1d (a)	5.9	curr	5.9	glob	0	big	5.4	6.0	7.7	7.6	7.6	fnd
Roszman1 (a)	3.6	curr	3.6	glob	0	big	2.3	0	5.0	6.1	6.1	fnd
ENSO (a)	3.1	curr	3.1	glob	0	init	4.5	0	6.3	6.5	6.5	fnd
MGH09 (h)	0	fnd	0	glob	0	big	2.7	0	0	0	0	curr
Thurber (h)	1.3	curr	1.3	glob	0	big	6.4	6.4	0	8.9	8.9	fnd
BoxBOD (h)	0	fnd	0	glob	0	big	0	0	7.9	0	0	fnd
Rat42 (h)	5.0	fnd	5.0	err	0	big	6.3	5.8	8.0	8.2	8.2	fnd
MGH10 (h)	0	fnd	0	glob	0	big	6.7	0	0	0	0	fnd
Eckerle4 (h)	0	fnd	0	big	0	big	4.6	0	0	0	0	fnd
Rat43 (h)	0	fnd	0	err	0	big	0	0	6.9	0	0	fnd
Bennett5 (h)	0	curr	0	glob	0	big	0	0	1.8	0	0	fnd

**Table 13** Results for three methods of Excel 2010 Solver (GRG2, Multistart, Evolutionary), the two methods of Calc 3.3 (DPES, SCO) and Gnumeric NLSolve from Start II initial values. Abbreviations: “curr” means current, “fnd” means found, “glob” means global, “err” means error, “init” means initial, “noimp” means can’t improve

Test (level)	GRG2		Multistart		Evolutionary		Calc 3.3		Gnumeric NLSolve
	LRE	Note	LRE	Note	LRE	Note	DPES	SCO	
Misra1a (l)	6.0	curr	6.0	glob	1.0	init	5.7	5.7	8.5
Chwirut2 (l)	4.7	curr	4.7	err	0	init	5.8	5.1	7.5
Chwirut1 (l)	5.0	curr	5.0	glob	0	init	6.0	5.9	6.8
Lanczos3 (l)	0	curr	0	err	0	big	0	0	0
Gauss1 (l)	5.0	curr	5.0	err	1.1	init	6.0	0	8.5
Gauss2 (l)	0	curr	0	err	0	big	0	0	8.3
DanWood (l)	4.4	curr	4.4	err	1.0	init	4.7	5.4	8.3
Misra1b (l)	6.1	curr	6.1	glob	0	big	0	0	7.5
Kirby2 (a)	1.9	curr	1.9	glob	0	big	4.8	0	5.5
Hahn1 (a)	0	err	0	big	0	big	3.9	0	2.5
Nelson (a)	0.9	fnd/ init	0.9	glob	0	init	3.5	3.4	5.2
MGH17 (a)	1.2	curr	1.2	err	0	init	1.5	0	2.9
Lanczos1 (a)	0	curr	0	err	0	big	0	0	0
Lanczos2 (a)	0	curr	0	err	0	big	0	0	0
Gauss3 (a)	3.9	curr	3.9	err	0	init	6.4	0	7.6
Misra1c (a)	4.5	curr	4.5	err	1.2	init	6.7	5.4	8.2
Misra1d (a)	4.3	curr	4.3	glob	1.5	init	6.0	5.4	7.8
Rozzman1 (a)	0.7	fnd	0.7	glob	0	init	2.4	2.1	7.2
ENSO (a)	3.1	curr	3.1	glob	0	init	1.2	1.2	6.3
MGH09 (h)	3.7	curr	3.7	glob	0	big	2.1	1.6	7.2
Thurber (h)	0	curr	0	glob	0	big	0	1.1	4.5
BoxBOD (h)	5.5	curr	5.5	glob	0	big	6.4	6.6	8.0
Rat42 (h)	4.9	curr	4.9	err	2.0	noimp	6.0	5.7	8.8
MGH10 (h)	0	curr	0	glob	0	big	6.6	0	0
Eckerle4 (h)	4.5	fnd	4.5	glob	0	init	4.1	4.4	9.0
Rat43 (h)	2.9	curr	2.9	err	1.3	init	5.8	2.1	6.0
Bennett5 (h)	0	curr	0	glob	0	big	0	0	0.8

Now, with Start II (close) initial values, the LRE's in Table 13 are often better than those in Table 12 for GRG2, but nevertheless 8 times with zero digit accuracy. Almiron et al. (2010) noticed that results with Start II are in a general matter better than those with Start I. This is confirmed but, in some cases, the results are surprisingly much worse with a loss of at least one digit (Misra1c, Roszman1) or even with zero digit accuracy (Gauss2, Thurber). In one case only (Hahn1), the GRG2 algorithm did not converge and issued an error. Note also that for another case (Nelson), the local optimum found was at the starting point. For the Multistart and the Evolutionary methods, the conclusions are similar to those in Table 12 even if there are some variations. For the last two cases mentioned (Hahn1 and Nelson), the final point was considered as a global solution. Except in one case (Rat43),  $LRE > 1$  for the Evolutionary method simply indicates that the initial values are close to the certified values.

These results for the two new methods seem so strange that the author contacted Frontline Systems (see above) to ask for their help and installed a free trial of Premium Solver Pro 11.0 in order to compare the results on some datasets. The tentative conclusions with Premium Solver Pro are (i) that the error message noticed 13 times out of 27 for Multistart does not appear and thus that an Answer report is issued, and (ii) that the behavior of the Evolutionary method is the same. We also noticed that Multistart is faster but Premium Solver Pro is advertised as being faster on some tasks so this may not be surprising. We also tried to find explanations. The answers that were given is that, in their quest for a global maximum, both the Multistart and the Evolutionary method make use of bounds on the variables and if no bound is provided, like in the StRD benchmark tests, they select bounds but these bounds are apparently much too big, leading to huge values. It is also explained that these methods are designed for complex problems, preferably with non smooth functions, and therefore may fail on problems with smooth functions. As a final conclusion, the statisticians should not hope that the new Excel 2010 Solver will improve their fits, at least on the light of these StRD benchmark datasets, at least without supplying bounds for the parameters. Nevertheless, as Billo (2012, personal communication) says, "If the Solver gives a good visual fit, then the coefficients are useful".

### 5.3 Calc Solver for Nonlinear Regression extension

Another contribution of Almiron et al. (2010) is, as it has already mentioned in Section 1, to use OpenOffice.org Calc and its customized siblings. Almiron et al. (2010) have noticed differences among setups for different hardware platforms (i386, amd64) and different operating systems (Windows, Ubuntu, MacOS). As said above, we have just considered a Windows operating system on an i386 platform. Calc integrated solver covers only linear programming, but an extension called Solver for Nonlinear Regression is available under a LGPL license. That solver contains two non-deterministic evolutionary algo-

rithms called DPES (differential evolution particle swarm) and SCO (social cognitive optimization). We have tried to reproduce the Almiron et al. (2010) results but using a convergence tolerance of  $10^{-7}$  like with Excel 2010 Solver. The results for these DPES and SCO methods are given in the last two columns of Table 12 and Table 13, and are close indeed. The methods sometimes take a long time (up to 20 minutes), provide no detailed report similar to Excel “Answer Report” and stopping is generally due to stagnation, i.e. no improvement. Looking at Start I initial values, the number of cases with zero digit accuracy is smaller (9 out of 27) for DPES, but higher (18) for SCO. Nevertheless in 14 cases the accuracy obtained by DPES is better than for Excel 2010 Solver with set 1 options although it is worse in 6 cases. SCO is not as good as DPES on the StRD datasets but when it works, LRE is greater than 5, which occurs rarely (4 times out of 27) for Excel 2010 Solver. The performance of DPES and of SCO is still slightly better from the Start II initial values although small LRE’s occur more frequently.

#### 5.4 Gnumeric NLSolve solver

We were suggested to add Gnumeric NLSolve to the analysis in Tables 11 and 12. This is sometimes tricky in Release 1.10.16, (i) due to messages saying that “Gnumeric is unable to locate the program glpsol needed for the Gnu Linear Programming Kit solver” and then “Gnumeric is unable to locate the program lp\_solve n”), (ii) the need to put all parameters in one range and specify nonlinearity. The options are also minimal (maximum number of iterations and time limit) and the same message always appeared saying that it was finished. But the results are sometimes very good even for Start I initial values. Except in one case (Thurber) it is as good as or better than Excel 2010 with set 1 options with zero digit accuracy in 9 cases. Again, aside from the Thurber test, the results are comparable (slightly better or slightly worse) to those of Excel 2010 with set 2 options which had 11 cases with zero digit accuracy. For Start II initial values, the numbers of digits obtained by Gnumeric are greater than those using Excel 2010 with set 1 options, generally by a few units, with the exception of 5 cases with zero digit accuracy (instead of 8).

#### 5.5 Conclusion for nonlinear regression

To conclude our experiments with Excel 2010 Solver for nonlinear regression, messages issued are more explanatory than in previous versions but in only one of the StRD problems (Rat42, with 5-digit accuracy) could GRG2 with set 1 options find the certified local optimum. Otherwise, either it pretended to have found a (zero digit) solution, or it stopped too early with the message “Solver has converged to the current solution”. With set 2 options, we got considerably better results, not only higher LRE but also better justified “fnd” messages. Moreover, the “fnd” message was issued in 6 cases where zero digit accuracy was obtained.

Surprisingly, with the Multistart option, a different message is issued, stating that a global optimum has been reached, which is of course completely wrong since not even a local optimum is attained.

Nevertheless, looking at the results with the two evolutionary algorithms in Calc Solver for Nonlinear Regression extension, which take much more time but are comparatively slightly better, there is still room for improvement in Excel 2010 Solver. This feeling is still stronger when Gnumeric NLSolve is considered since it is very successful, even without the opportunity of specifying options. It is a pity that, after more than ten years of papers in the statistical literature, no remedy has been found to repair Excel failures in nonlinear regression.

## 6 Data graphical representation

### 6.1 Generalities

The beginning of Section 20 of Heiser (2009) website is a good summary on how bad Excel was for plotting statistical data, see also Su (2008). Let us summarize Su (2008)'s findings:

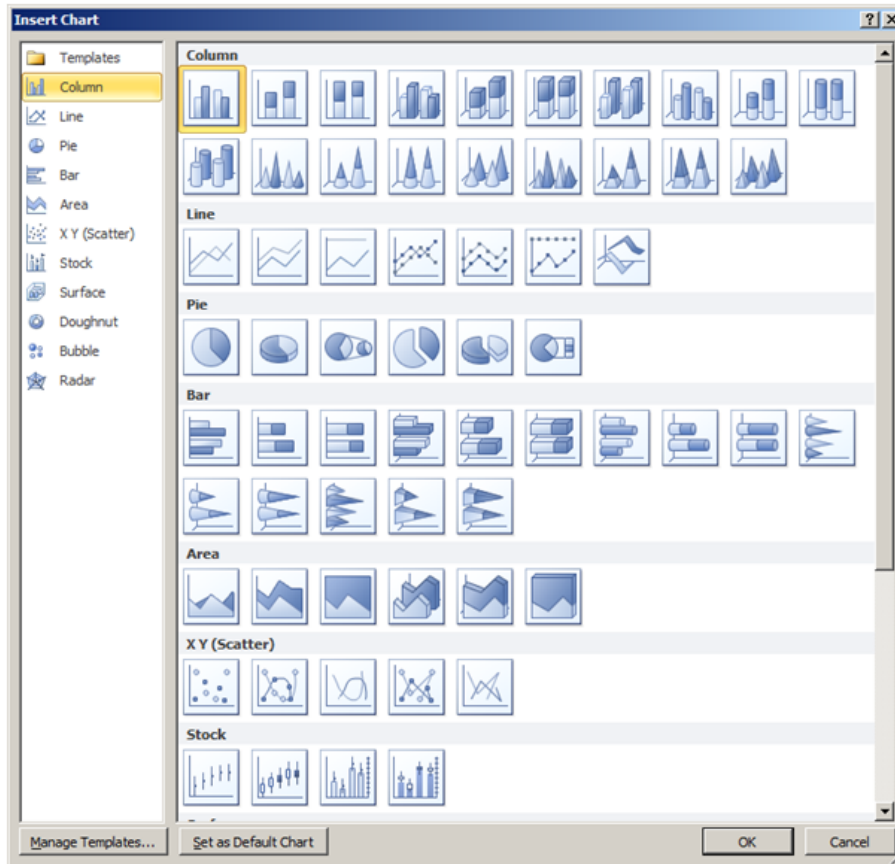
- default scattergrams display inappropriate  $y$ -axis, superfluous (or even wrong) legend and labels, and too many axis labels which can better be replaced by ticks, but hide data points because default filled diamonds are too big;
- junk charts are easily produced by a few clicks, like those with generally useless 3-D effects;
- in general, users have to expend great effort to clean up the chartjunk in Excel defaults.

Some examples of badly designed line, pie, area, bubble and bar charts that violate the principles of statistical graphics are shown. In each case revised and improved charts are proposed. Making effective graphs is not easy (Tufte (1983), Cleveland (1994)) and Excel does not facilitate it, see Walkenbach (2007).

It should be noted that the quality of the adjustments made on the basis of plots has also been criticized, see Hargreaves and McWilliams (2010). If we keep only problems that were not corrected in subsequent service packs, they have shown

- that equations of polynomial fits added to a plot can be badly formatted;
- that the computation of  $R^2$  when the forced-intercept trend line is not consistent;
- that nonsense trend lines can easily be added for categorical  $X$  variables.

Fig. 5 Suggestion of plots in Excel 2010 (excerpt)



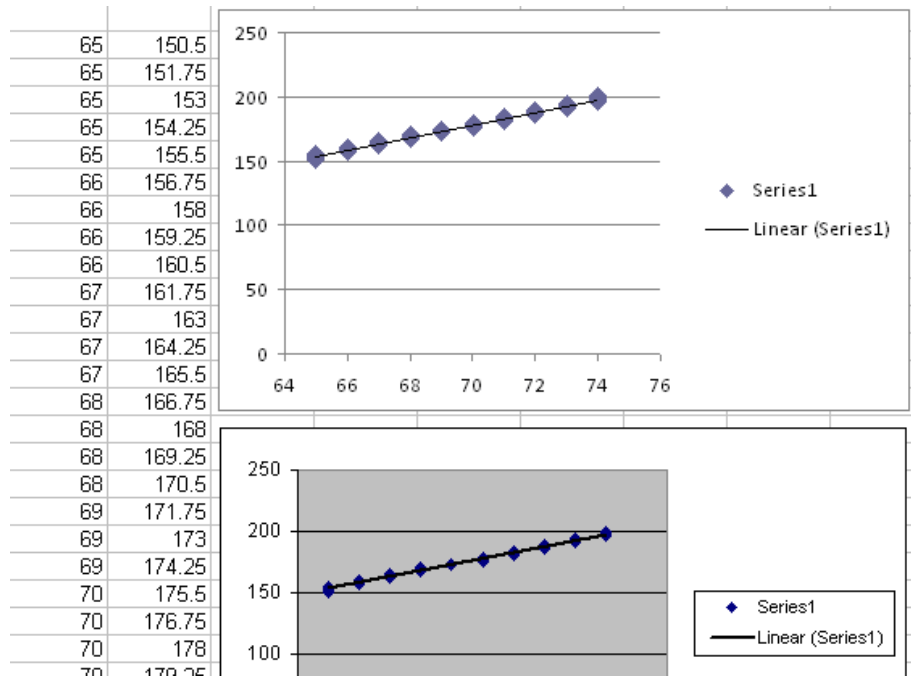
## 6.2 Excel 2010 plots

Although Excel 2010 brings some improvements (the maximum number of points being plotted is increased, formatting is easier, and it is now possible to record macros while preparing plots), there is no improvement in Excel 2010 on the capability of producing statistical plots, see Fig. 5. Stock plots can only be used as a rough approximation of boxplots. Therefore most, if not all, of the criticisms made by Su (2008) are still valid. Indeed Fig. 6 shows the analogue of Su (2008)'s Fig. 1 obtained using Excel 2010. Note that in the (linear) regression statistical add-in, the "normal probability plot" is still bearing that name despite it not being a normal probability plot. See McCullough and Heiser (2008) but the appearance is better than in Excel 2007.

The remarks of Hargreaves and McWilliams (2010) on fits are still valid for Excel 2010 but that paper was probably too recent to have been taken into account by Microsoft developers.



**Fig. 6** Replication of Su (2008)'s Fig. 1 in Excel 2010 (top) and Excel 2003 (bottom) with our own data



### 6.3 Calc plots

OpenOffice.org Calc 3.3 does not allow to produce chartjunk as easily as Excel, but there is no way of producing statistical plots like box plots. The equivalent of Fig. 4 of Su (2008) is not better. Again, some (free) extensions allow the producing of statistical plots as well as statistical add-ins. With regard to trend lines aspects criticized by Hargreaves and McWilliams (2010), forced-intercept fits are not supported and the formatting trend line equations are accurate (even with too many digits), but it is possible to add a trend line to a categorical column chart.

### 6.4 Gnumeric plots

Gnumeric 1.10.6 has a few statistical plots: horizontal and vertical boxplots with or without visible outliers and two types of histograms.

## 6.5 Conclusion for data graphical plots

Producing statistical plots is not the strongest point of spreadsheet processors, with Gnumeric as the leader among the software packages examined here.

## 7 General conclusions

This paper has clarified Microsoft's claims that they improved the statistical functions included in Excel 2010, the random number generator and the Solver. Some of the functions themselves (including those for computing probabilities and quantiles and generating pseudo-random numbers), but not all, have been changed and are indeed much better. The random number generator, more precisely the `RAND` function, is also better but note that neither the random numbers in the Data Analysis Toolpak, nor the VBA `rnd` function have changed. More generally the Data Analysis Toolpak has not been changed at all. The new Solver has two new methods that will not improve statistical fits, except perhaps if suitable bounds can be set to the parameters. The old GRG2 algorithm is unfortunately not better for nonlinear regression and still too often returns results with zero digit accuracy. Reports are slightly clearer but there is a persistent need to improve stopping criteria, default tolerance settings or messages. Finally, the statistical aspects in graphs are not better.

We believe that, without a prompt reaction from Microsoft, the already very critical view against using spreadsheets for doing any statistical analysis, like expressed in Heiser (2009) and IBM (2010), will be even more justified. It is standard practice for software developers to fix errors quickly and correctly. Microsoft has fixed the errors in the statistical procedures of Excel neither quickly nor correctly. The recent improvements reported in this paper should not hide the fact that Microsoft is still marketing a product that contains known errors. We didn't analyze Excel in Office 2013 but, according to Microsoft (2013), where the changes with respect to Office 2010 are collected, there are few changes to Excel and nothing about the statistical aspects is mentioned.

One referee suggested that we can cite the Reinhart-Rogoff vs Herndon-Ash-Pollin debate (Herndon et al. , 2013, Reinhart and Rogoff, 2010). That debate only happened partly due to a mistake in the use of an Excel function, not due to a problem in the functions themselves. Clearly, bad uses of spreadsheet functions is out of the scope of this paper. Interested readers can have a look at the website of the European Spreadsheet Risks Interest Group (2013) for other spreadsheet horror stories.

This paper has also updated some comments on OpenOffice.org Calc contained in Yalta (2008) and Almiron et al. (2010). Of course users have more incentives in suggesting changes in open source software. For example we were pleased to see that a Mersenne Twister algorithm appeared as a Calc extension

due to our investigations. McCullough (2004c) had made a similar observation for Gnumeric.

## Acknowledgements

I would like to thank Christian Ritter, Pierre Dagnelie, Bruce McCullough, and Talha Yalta for some references and Atika Cohen and two anonymous referees of a first version for their comments. I am deeply grateful to Bruce McCullough and Talha Yalta who kindly gave me their test files, for their support and their comments. I am not sure I would have completed this paper without the insistence of Bruce McCullough. I thank Richard Simard for his advice on TestU01. I would like to thank David Heiser, Daniel Fylstra and Skylab Gupta, both of Frontline Systems, for their comments on a second version and Joseph Billo on a third version. I am indebted to Jerry W. Lewis for his remarks, although not all could be addressed. I would like to thank the two anonymous referees of the last version of the paper and Wassim Cornet for editing. I assume full responsibility for the paper. It is dedicated to the memory of Leo Knüsel who passed away on April 22, 2013.

## References

- Almiron MG, Lopes B, Oliveira ALC, Medeiros AC, Frery AC (2010) On the numerical accuracy of spreadsheets. *J Statist Softw* 34 (4):1-29.
- Bailey DH (1993) A portable high performance multiprecision package. RNR Tech. Rep. 90-022, NASA Ames Research Center, Moffett Field, Calif.
- Berger RL (2007) Nonstandard operator precedence in Excel. *Comput Statist Data An* 51:2788-2791.
- Cleveland W (1994) *The elements of graphing data*, 2nd edn. Hobart Press, Summit, NJ.
- Crabtree C (2009) How random is random. [http://www.cliffcrabtree.com/index.php?option=com\\_content&view](http://www.cliffcrabtree.com/index.php?option=com_content&view). Accessed 7 September 2010.
- Cryer JD (2001) Problems with using Microsoft Excel for statistics. Proceedings of the Joint Statistical Meetings, American Statistical Association, Atlanta.
- European Spreadsheet Risks Interest Group (2013) <http://www.eusprig.org/index.htm>. Accessed 15 November 2013.
- Fylstra D, Lasdon L, Watson J, Waren AD (1998) Design and use of the Microsoft Excel Solver. *Interfaces* 28:29-55.
- Guibas LJ, Odlyzko AM (1981) Periods in strings. *J Combin Theory A* 30:1942.
- Hargreaves BR, McWilliams TP (2010) Polynomial trendline function flaws in Microsoft Excel. *Comput Statist Data An* 54:1190-1196.

- Heiser DA (2009) Microsoft Excel 2000, 2003 and 2007 faults, problems, workarounds and fixes. <http://www.daheiser.info/excel/frontpage.html>. Accessed 26 February 2013.
- Herndon T, Ash M, Pollin R (2013) Does high public debt consistently stifle economic growth? A critique of Reinhart and Rogoff, PERI working paper n 322, April, University of Massachusetts Amherst.
- Hümmer T (2010) International OpenOffice market shares. <http://www.webmasterpro.de/portal/news/2010/02/05/international-openoffice-market-shares.html>. Accessed 26 February 2013.
- IBM (2010) The risks of using spreadsheets for statistical analysis. IBM Software Business Analytics, Somers, <http://public.dhe.ibm.com/common/ssi/ecm/en/ytw03240usen/YTW03240USEN.PDF>. Accessed 28 February 2013.
- Keeling KB, Pavur RJ (2011) Statistical accuracy of spreadsheet software. *Am Statist* 65:265-273.
- Knüsel L (1998) On the accuracy of statistical distributions in Microsoft Excel 97. *Comput Statist Data An* 26:375-377.
- Knüsel L (2003) Computation of statistical distributions - documentation of program ELV, 2nd edition. <http://www.statistik.lmu.de/~knuesel>. Accessed 26 February 2013.
- Knüsel L (2005) On the accuracy of statistical distributions in Microsoft Excel 2003. *Comput Statist Data An* 48:445-449.
- Knüsel L (2011) On the accuracy of statistical distributions in Microsoft Excel 2010. [http://www.csdassn.org/software\\_reports/Excel2011.pdf](http://www.csdassn.org/software_reports/Excel2011.pdf). Accessed 26 February 2013.
- Koenker R (2005) Quantile regression. Cambridge University Press, Cambridge.
- Lasdon LS, Waren AD, Jain A, Ratner M. (1978) Design and testing of a generalized reduced gradient code for nonlinear programming. *ACM T Math Softw* 4:34-50.
- L'Ecuyer P, Simard R (2007) TESTU01: A C library for empirical testing of random number generators. *ACM T Math Softw* 33 (22):1-40.
- L'Ecuyer P, Simard R (2009) TestU01: A software library in ANSI C for empirical testing of random number generators. User's guide, compact version. <http://www.iro.montreal.ca/~lecuyer/>.
- McCullough BD (1998) Assessing the reliability of statistical software: part I. *Am Statist* 52:358-366.
- McCullough BD (1999) Assessing the reliability of statistical software: part II. *Am Statist* 53:149-159.
- McCullough BD (2000) The accuracy of Mathematica 4 as a statistical package. *Comput Statist* 15:279-299.
- McCullough BD (2004a) Some details of nonlinear estimation. In Altman M, Gill J, McDonald MP (ed) *Numerical issues in statistical computing for the social scientist*, Chapter 8. Wiley, New York, pp 199-218.
- McCullough BD (2004b) Wilkinson's tests and econometric software. *J Econ Soc Meas* 29:261-270.

- McCullough BD (2004c) Fixing statistical errors in spreadsheet software: the cases of Gnumeric and Excel. CSDA Statistical Software Newsletter. [www.csdassn.org/software\\_reports.cfm](http://www.csdassn.org/software_reports.cfm).
- McCullough BD (2006) A review of TESTU01. *J Appl Econom* 21:677-682.
- McCullough BD (2008a) Editorial: special section on Microsoft Excel 2007. *Comput Statist Data An* 52:4568-4569.
- McCullough BD (2008b) Microsoft Excel 2007's 'Not the Wichmann-Hill' random number generator. *Comput Statist Data An* 52:4587-4593.
- McCullough BD, Heiser DA (2008) On the accuracy of statistical procedures in Microsoft Excel 2007. *Comput Statist Data An* 52:4570-4578.
- McCullough BD, Renfro CR (2000) Some numerical aspects of nonlinear estimation. *J Econ Soc Meas* 26:63-77.
- McCullough BD, Wilson B (1999) On the accuracy of statistical procedures in Microsoft Excel 97. *Comput Statist Data An* 31:27-37.
- McCullough BD, Wilson B (2002) On the accuracy of statistical procedures in Microsoft Excel 2000 and Excel XP. *Comput Statist Data An* 40:713-721.
- McCullough BD, Wilson B (2005) On the accuracy of statistical procedures in Microsoft Excel 2003. *Comput Statist Data An* 49:1244-1252.
- McCullough BD, Yalta AT (2013) Spreadsheets in the cloud - Not ready yet. *J Statist Softw*, 52 (7):1-14.
- Matsumoto M, Nishimura T (1998) Mersenne Twister: a 623-dimensionally equidistributed uniform pseudorandom number generator. *ACM T Model Comput Simul* 8:3-30.
- Microsoft Office Blog (2009a) Function improvements in Microsoft Excel 2010, White paper. [http://blogs.office.com/cfs-file.ashx/\\_key/CommunityServer-Components-PostAttachments/00-00-00-49-53/ExcelFunctionImprovements-10\\_2D00\\_05\\_2D00\\_09\\_-2800\\_1\\_2900\\_.pdf](http://blogs.office.com/cfs-file.ashx/_key/CommunityServer-Components-PostAttachments/00-00-00-49-53/ExcelFunctionImprovements-10_2D00_05_2D00_09_-2800_1_2900_.pdf). Accessed 26 February 2013.
- Microsoft Office (2009b) About Solver. <http://office.microsoft.com/en-us/excel-help/about-solver-HP005198368.aspx?pid=CH010004571033>. Accessed 26 February 2013.
- Microsoft TechNet (2013) Changes in Office 2013 <http://technet.microsoft.com/en-us/library/cc178954.aspx>. Accessed 11 December 2013.
- Oppenheimer DM (2009a) Function improvements in Excel 2010. <http://blogs.msdn.com/b/excel/archive/2009/09/10/function-improvements-in-excel-2010.aspx>. Accessed 26 February 2013.
- Oppenheimer DM (2009b) New and improved solver. <http://blogs.office.com/b/microsoft-excel/archive/2009/09/21/new-and-improved-solver.aspx>. Accessed 26 February 2013.
- Reinhart CM, Rogoff KS (2010) Growth in a time of debt. *Am Econ Rev* 100:573-578.
- Rinnooy-Kan AHG, Timmer GT (1987a) Stochastic global optimization methods. Part I: clustering methods. *Math Program* 39:27-56.
- Rinnooy-Kan AHG, Timmer GT (1987b) Stochastic global optimization methods. Part II: multi level methods. *Math Program* 39:57-78.

- Sawitzki G (1994) Report on the numerical reliability of data analysis systems. *Comput Statist Data An* 18:289-301.
- Simon G (2000) Excel produces numerically inaccurate results. <https://www.jiscmail.ac.uk/cgi-bin/wa.exe?A2=ind0012&L=ASSUME&F=&S=&X=31BDFD6E4E1E0E4F11&P=4325>. Accessed 5 September 2010.
- Su Y-S (2008) It's easy to produce chartjunk using Microsoft Excel 2007 but hard to make good graphs. *Comput Statist Data An* 52:4594-4601.
- Tufte ER (1983) *The visual display of quantitative information*. Graphic Press, Cheshire.
- Ueberhuber CW (1946) *Numerical computation: methods, software, and analysis*. Springer, Berlin.
- Walkenbach J (2007) *Excel 2007 Charts*. Wiley, New York.
- Wichmann BA, Hill ID (1982) Algorithm AS 183: An efficient and portable pseudo-random number generator. *J Roy Statist Soc Ser C Appl Statist* 31:188-190. Reprinted with correction in Hill ID and Griffith P (eds) *Applied statistics algorithms*. Ellis Horwood, Chichester, 1985.
- Yalta AT (2008) The accuracy of statistical distributions in Microsoft Excel 2007. *Comput Statist Data An* 52:4579-4586.
- Yalta AT and Yalta AY (2010) Should economists use open source software for doing research? *Comput Econom* 35:371-394.
- Yalta AT and Jenal O (2009) On the importance of verifying forecasting results. *Int J Forecasting* 25:62-73.