On the accuracy of statistical procedures in Microsoft Excel 2010

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.

Abstract

All previous versions of Microsoft Excel have been criticized by statisticians for several reasons, including the accuracy of statistical functions, the properties of random number generator, the quality of statistical add-ins, the weakness of the Solver for nonlinear regression, and the data graphical representation. 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.

Keywords:
statistical function, statistical procedure, random number generator, nonlinear regression, Excel 2010, OpenOffice.org Cal 3.3, Gnumeric

---

1. Introduction

1.1. Generalities

Criticisms on the statistical aspects of the various versions of Microsoft Excel have 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. A first (timid) improvement appeared in Office 2003 with a better handling of extreme observations, more robust statistical functions, a different pseudo-random number generator. 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).

The most recent contributions to the subject are a whole section in 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). In these 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 these 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 disagreement by statisticians:

1. Accuracy of the statistical functions. Yalta (2008) has shown that Excel 2007 is better than Excel 2003 for several 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.

2

Yalta and Jenal (2009).

4. The Solver module. It is used in nonlinear fits. It often claims erroneously 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

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

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, is 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.

1.2. Excel 2010

Released on May 12, 2010 officially, Office 2010 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 (2009). For the Solver module, we only consider nonlinear optimization without constraints since the other aspects would require more operational research skills. 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 some of the other aspects.

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 released. 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 have decided to develop a LibreOffice suite outside of Oracle.

1.4. Gnumeric 1.10.16

It is known ((McCullough , 2004c), (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) who had discovered some errors in a previous version has reported that "The few part-time volunteers who maintains and develop Gnumeric fixed all the problems in a few weeks."

2. Accuracy of the statistical functions

2.1. Generalities

As indicated above, the most recent contribution on this subject is Yalta (2008). Through a series of tables, each devoted to a specific function for

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 difference 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	%Excel-Calc=0
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-9	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-5	1E-13	66
BETAINV	8	14/36+24	14.0	9E-15	9E-14	85
TINV #	9	15/72+8	5.4	3E-6	3E-6	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	2E-4	2E-4	73
BETADIST	-	0/156	13.4	3E-14	1E-14	74

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, still worse, 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

5

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

have 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 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 contents is put instead of LRE. Cases entries which were added with respect to Yalta (2008) are shaded.

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 *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 have also added values obtained from Excel 2010 and Calc 3.3 and sent the updated file to his original author. A synthesis of the functions treated and number of cases in Tables 2 to 10 and in the

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

bigger original file is shown in Table 1. We added also 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 4: Simple (if sigma is FALSE) or cumulative (if sigma is TRUE) Poisson probabilities at k for different values of mean parameter λ

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

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 shown 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.

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 3.0whose errors apparently prevented to save the workbook.

2.5. Gamma distribution (Yalta's Table 5)

For the gamma distribution shown in Table 5, Excel 2010 appears 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.

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

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.

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

As regards 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. Then weakest results are for very small probabilities and no longer large degrees of freedom.

2.8. Beta distribution (Yalta's Table 8)

About the specific quantiles of the beta distribution depicted in Table 8, Excel 2010 is really 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.

2.9. Student distribution (Yalta's Table 9)

For the tested quantiles of a Student distribution, Table 9 reveals that Excel 2010 is really 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.

9

Table 6: Quantile of order p of a standard normal distribution

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

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

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 χ^2 and Student quantiles that Excel 2010 could compute whereas Excel is not accurate enough for some χ^2 quantiles. By the way, fractional degrees of freedom are still truncated to integers.

Note the addition of new functions in Excel 2010 with more explicit names. For more details, see Microsoft (2009). 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, by the way) since the

Table 7: Quantile of order $1-p$ of a χ^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

new names will not be recognized. For reasons of compatibility, Microsoft has hopefully maintained the old names (which we employed here, mentioning also corresponding new names) while improving accuracy. Be careful that CHISQ.DIST, T.DIST and F.DIST now refer to cumulative distribution functions, contrarily to CHIDIST, TDIST and FDIST, and that CHISQ.INV, T.INV and F.INV now refer to quantiles for given probabilities, contrarily to CHIINV, TINV and FINV.

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-784 arithmetic, although some algorithms have been used to improve the calculations, like TWOSUM to computing the rounding error when adding two numbers, see Microsoft (2009).

These results have been confirmed by a check in the more detailed file from which Yalta (2008) shows 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 ELV Ed. 2 (Knüsel , 2003), on the other hand, but of course only when

11

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

ELV values were provided, thus excluding e.g. probabilities smaller that 1.E-100. It can be seen the order of magnitude is smaller than 5.E-6, meaning that 5 or 6 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 χ^2 quantiles are not taken into account correctly (although limiting distribution arguments can be used in such a case) and this indicates unsufficient testing. Fourth, past track of Microsoft failures in the domain of statistical functions in 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. Overall functions and test cases, the smallest non zero difference is -7.2910^{-301}. Since it is known that even the two associative laws and

Table 9: Quantile of order $1 - p/2$ of a Student distribution with 1 degree 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!

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.

Moreover, 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. Let us now look if the improvements also occur in the other areas. See again Microsoft (2009) for the declared improvements.

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:

Table 10: Quantile of order $1 - p$ of a Fisher-Snedecor distribution with 1 and 1 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

- (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 menu Tools > Data Analysis > Random Number Generation in Excel 2003;

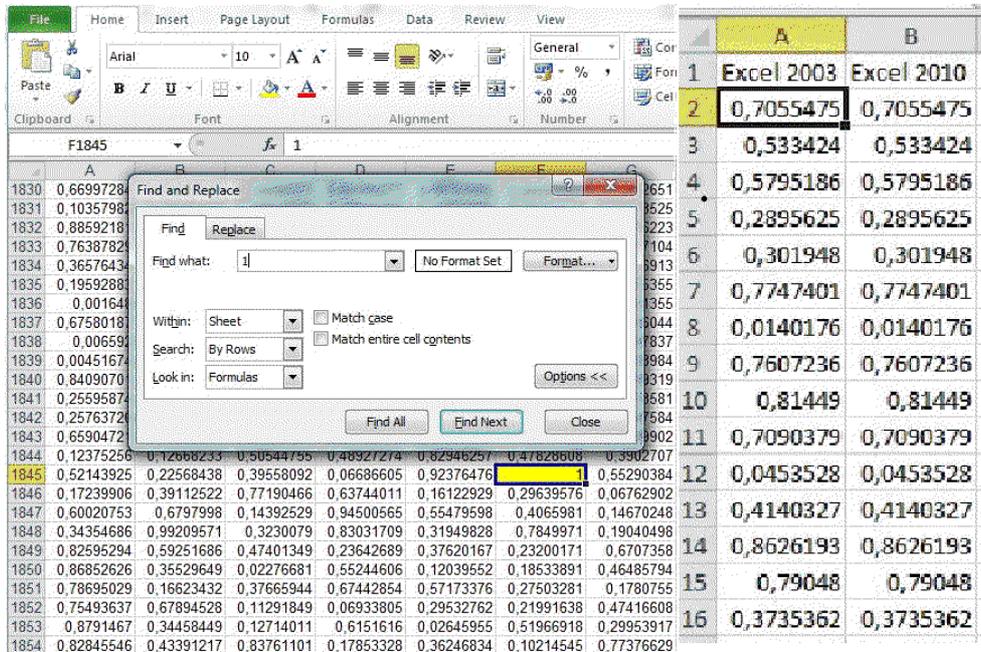- (C) the generator used in VBA `rnd` function.

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, and recalled that the period of the generator is $6.95 \ 10^{12}$, 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 provide 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.

3.2. Excel 2010

Microsoft (2010a) 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 Figure 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.

Figure 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 2007 and 2010

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 checked empirically (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 2007. See the right hand side part of Figure 1. Hence, there is no improvement in this area either.

Figure 2: Result of hexadecimal conversion of a sequence of pseudo-random numbers obtained in Calc 3.3 using `RAND` function
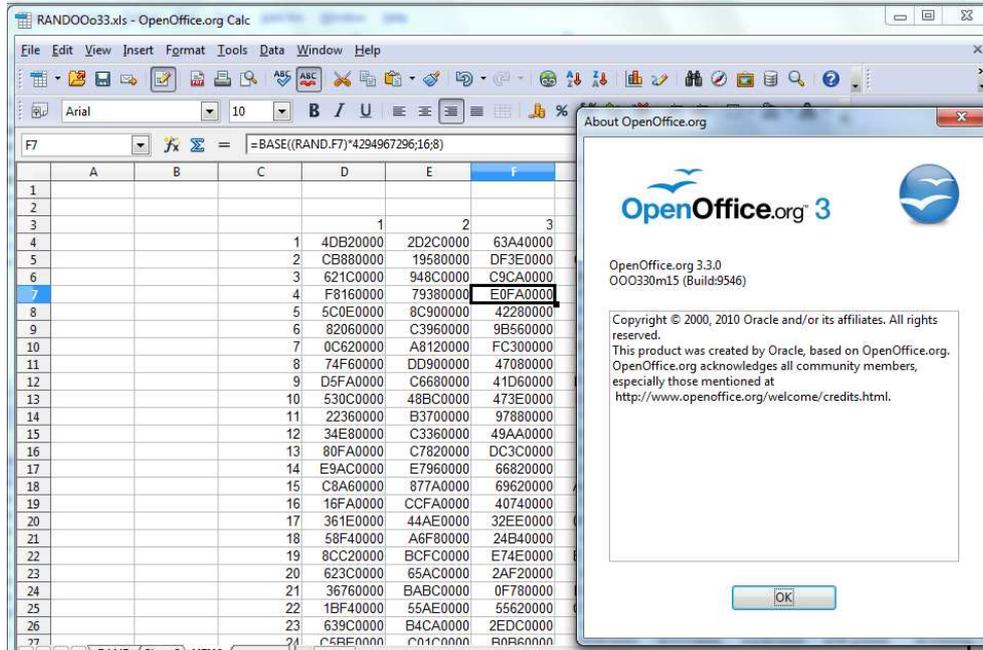
Let us now check the new generator (A). We applied a procedure similar to that proposed by McCullough (2008b), sections 3 and 4, who performed high-precision calculations with the Python language. Here, we carried out our calculations with a FORTRAN program making use of a multiprecision library (Bailey , 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 Microsoft (2010c), is that the new generator makes use of the Mersenne Twister algorithm, Matsumoto and Nishimura (1998). Such a change is in agreement with suggestions of several authors, e.g. McCullough and Wilson (2005), McCullough (2008b), to provide a generator with a longer period and with better properties. That generator passes most tests in L'Ecuyer and Simard (2007) and its period is $2^{19937} - 1$. L'Ecuyer and Simard (2007) mention that it passes all Small-Crush tests, and all but two tests of the bigger batteries of empirical tests

16

called Crush and BigCrush. Simard (2011) has mentioned that the two test 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 32×1024 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 Small-CrushFile 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 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.

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). Unfortunately, this is not possible but there are external solutions.

3.3. Calc 3.3

To our knowledge, except 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 on a computer equiped with Windows XP and OpenOffice.org Calc 3.3 the experiment of Crabtree (2009), which was performed on an unspecified version of Calc. The

observations reproduced in Figure 2 are identical. Indeed, after multiplication by 2^{32}, the numbers seen in hexadecimal representation end up all with "0000", showing only 32-bit accuracy. According to messages in forums, it is known that the maximum number that can be reached is 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 depends 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 `stdlih.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 can reach 0.9999999995343.

We made some experiments (generation of random numbers in a 2048 × 128 array) that indeed confirm these suspected differences. This could be checked in a Mandriva 2009 Linux installation using Calc 3.1 and on an Ubuntu 10.10 installation using Calc 3.2 (where we got 0.999999694061 as maximum, thus greater than 0.9999695). Of course, the corresponding analogue of Figure 2 shows hexadecimal numbers that do not end with "0000".

There does not seem an intention to improve the random number generator but the advantage of a free source software is that, in principle, anybody could do it for his/her own usage and making it profitable to other people. During our investigations, an extension MTRNG of Calc appeared, due to contributor Hanya, with a Mersenne Twister algorithm, Matsumoto and Nishimura (1998). The only problem is that there are lots of extensions of OpenOffice.org and only informed users will take care of that one.

3.4. Gnumeric 1.10.16

Gnumeric makes use of the same Mersenne Twister algorithm for several years, apparently because of McCullough (2004c) remarks.

4. Quality of the statistical add-ins

4.1. Generalities

The statistical add-ins were available by menu 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 $X_i = i$, $i = 1, ..., 10$, 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 define 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.

4.2. Excel 2010

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

- 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.

19

Figure 3: Regression by Excel 2010 on Wilkinson IV-C test

Task IV.C					
SUMMARY OUTPUT					
Regression Statistics					
Multiple R	1				
R Square	1				
Adjusted R Square	0.857142857				
Standard Error	6.1672E-16				
Observations	9				
ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	60	30	1.58E+32	6.87763E-96
Residual	7	2.66241E-30	3.8E-31		
Total	9	60			
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*
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

- Wilkinson's Test IV-C, which Microsoft Excel 4.0 failed, as described in Sawitzki (1994), meaning that a regression of a constant on some two linearly dependent variables is accepted.

This last point is worth discussion. 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 Figure 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 ses. 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 $-99999990+$ 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. We have

20

checked (see details in Mélard (2011)) that the other feasible (some are not feasible in Excel like tests IIB or IIIA) Wilkinson's tests are passed.

Since the other statistical add-ins have apparently not been changed, we have not checked all the StRD datasets. We have tried NumAcc2, NumAcc3, and NumAcc4 for the standard deviation and lag 1 autocorrelation, and SmLs07, SmLs08, and SmLs09 for ANOVA models. We reach conclusions similar to previous studies.

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. Conclusion

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 11 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 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, with also 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 there 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 a 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 recommend 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 behavior 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 a 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).

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) that this message occurs if the L_∞ norm of the objective gradient is less or equal 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

23

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 including "Solver cannot improve the current solution" ("noimp" in Tables 11 and 12).

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 recommendation "Try a smaller convergence setting or a different starting point". Also, there are more details in the reports about the options, see Figure 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". 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. 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).

Figure 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 0.0000001, 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
G2	B2 RSS	186382.3817	9771.499976

Variable Cells

Cell	Name	Original Value	Final Value	Integer
C3	B1	1	172.5000001	Contin
D3	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 intended not at finding 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

25

found a global optimum", denoted "glob" in Tables 11 and 12.

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 Table 11 and 12. We have used Newton search default only, since the conjugate search option was systematically inferior in the results of Almiron et al. (2010). Table 11 and Table 12 refer, respectively, to Start I (far) and to Start II (close) initial values. Columns 2 (GRG2), 4 (Multistart) and 6 (Evolutionary) provide problem LREs for these three methods, respectively. Columns 3 and 5 contain either a shortened form of GRG2 Solver message, "fnd" meaning that a local optimum was *found*, "curr" meaning that the Solver has converged to the *current* solution for the last five iterations, "glob" meaning that a *global* optimum was found, or "err" if the Solver has issued an *error* message preventing even the production of an Answer report, or "big" if a solution is given but showing at least one huge number. 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, typically of order 10^{30}, or "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 a set 2 of method parameters suggested by Fylstra and Gupta (2011) 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 11. 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 $>= 6$ 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).

Table 11: 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
Misra1a (l)	3.5	curr	3.5	err	0	big	0	5.2	8.7	8.7	fnd
Chwirut2 (l)	4.3	curr	4.3	err	0	init	5.4	5.6	7.5	8.1	fnd
Chwirut1 (l)	4.4	curr	4.4	err	0	init	5.7	6.6	8.1	8.1	fnd
Lanczos3 (l)	0	curr	0	err	0	big	0	0	0	0	curr
Gauss1 (l)	4.9	curr	4.9	err	0	init	6.6	6.8	8.6	9.2	fnd
Gauss2 (l)	4.6	curr	4.6	err	0	init	6.2	0	8.4	9.8	fnd
DanWood (l)	5.1	curr	5.1	glob	0	big	4.6	5.5	8.7	9.4	fnd
Misra1b (l)	3.5	curr	3.5	glob	0	big	0	0	7.8	8.3	fnd
Kirby2 (a)	1.1	curr	1.1	glob	0	big	5.1	0	5.4	6.6	fnd
Hahn1 (a)	0	curr	0	glob	0	big	3.8	0	0	0	*
Nelson (a)	0	fnd	0	err	0	noimp	0	0	5.1	7.4	curr
MGH17 (a)	0	fnd	0	err	0	big	2.0	0	0	0	fnd
Lanczos1 (a)	0	curr	0	err	0	big	0	0	0	0	curr
Lanczos2 (a)	0	curr	0	err	0	big	0	0	0	0	curr
Gauss3 (a)	4.0	curr	4.0	err	0	init	6.1	0	8.1	9.3	fnd
Misra1c (a)	6.3	curr	6.3	glob	0	big	5.6	5.8	8.2	7.8	fnd
Misra1d (a)	5.9	curr	5.9	glob	0	big	5.4	6.0	7.7	7.6	fnd
Roszman1 (a)	3.6	curr	3.6	glob	0	big	2.3	0	5.0	6.1	fnd
ENSO (a)	3.1	curr	3.1	glob	0	init	4.5	0	6.3	6.5	fnd
MGH09 (h)	0	fnd	0	glob	0	big	2.7	0	0	0	curr
Thurber (h)	1.3	curr	1.3	glob	0	big	6.4	6.4	0	8.9	fnd
BoxBOD (h)	0	fnd	0	glob	0	big	0	0	7.9	0	fnd
Rat42 (h)	5.0	fnd	5.0	err	0	big	6.3	5.8	8.0	8.2	fnd
MGH10 (h)	0	fnd	0	glob	0	big	6.7	0	0	0	fnd
Eckerle4 (h)	0	fnd	0	big	0	big	4.6	0	0	0	fnd
Rat43 (h)	0	fnd	0	err	0	big	0	0	6.9	0	fnd
Bennett5 (h)	0	curr	0	glob	0	big	0	0	1.8	0	fnd

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) while nevertheless showing the same solution as the GRG2 method, thus without improvement and detailed "Answer report" output. But also in all but one of the remaining cases (13 cases out of 27)

Table 12: 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
	LRE	Note	LRE	Note	LRE	Note	DPES	SCO	NLsolve
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
Roszman1 (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

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 saying that the maximum computing time was used appeared 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 case at

least one huge number is obtained in the proposed solution. Given that the Evolutionary takes stills more time (with a median of 50 s) than the Multistart method and never could find a solution, we can safely recommend *not* to use it.

Now, with Start II (close) initial values, the LRE's in Table 2 are often better than those in Table 1 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 1 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 (ii) 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.

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 a 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 algorithms called DEPS (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 11 and Table 12 and are close indeed. The methods take sometimes long (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, 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"), the need to put all parameters in one range and specify nonlinearity. Also the options are minimal (maximum number of iterations and time limit) and the same message always appeared telling 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 or better than Excel 2010 with set 1 options with zero digit accuracy in 9 cases. Except again 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

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 more well justified "fnd" messages but these were suggested by an expert and it is not clear that an expert can always help every user. 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 that a global optimum is 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 clear Excel failures in nonlinear regression.

6. Data graphical representation

6.1. Generalities

The beginning of Section 20 of Heiser (2009) web site 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.

Figure 5: Suggestion of plots in Excel 2010 (excerpt)
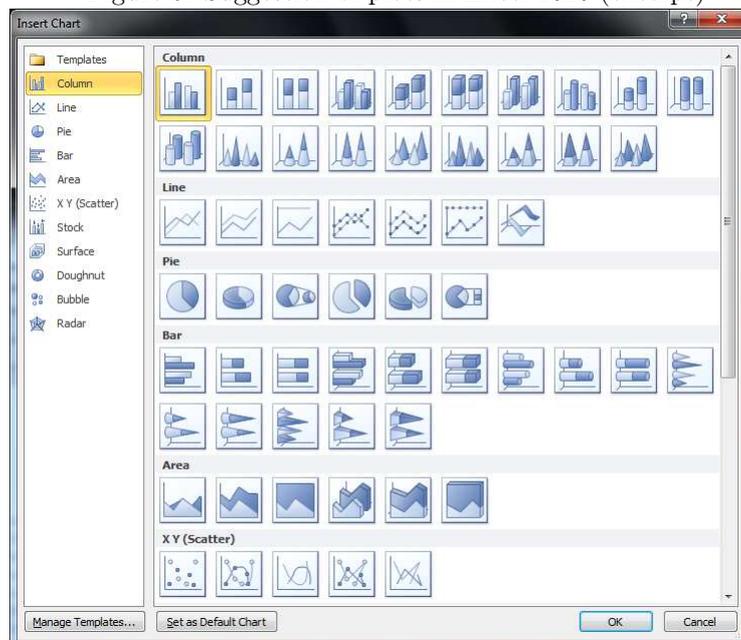
It should be noted that the quality of the adjustments made on the basis of plots has also been criticized, see Hargreaves and McWilliams (2010). If we keep only problems that were not corrected in subsequent service packs, they have shown

- that equations of polynomial fits added to a plot can be badly formatted;

- that that computation of R^2 when the forced-intercept trendline is not consistent;

- that nonsense trend lines can easily be added for categorical X variables.

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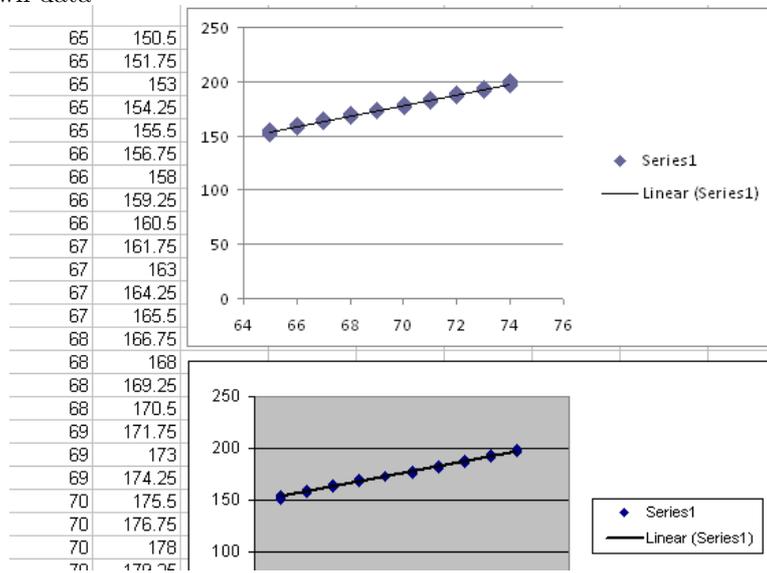
6.2. Excel 2010 plots

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 Figure 5. Therefore most, if not all, of the criticisms made by Su (2008) are still valid. Indeed Figure 6 shows the analogue of Su (2008)'s Fig. 1 obtained using Excel 2010. Note that the (linear) regression statistical add-in "normal probability plot" is still bearing that name despite it is not a normal probability plot. See

33

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 to recent to have been taken into account by Microsoft developers.

6.3. Calc and Gnumeric 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 Figure 4 of Su (2008) is not better. Again, some (free) extensions allow to produce statistical plots as well as statistical add-ins. About trendlines aspects criticized by Hargreaves and McWilliams (2010), forced-intercept fits are not supported and the formatting trendline equations are accurate (even with too many digits), but it is possible adding a trendline to a categorical column chart.

7. Conclusions

This paper has clarified Microsoft claims to have improved the statistical functions included in Excel 2010 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 but the Data Analysis Toolpak has not been changed and the statistical aspects in graphs are not better. The new Solver has two new methods that will not improve statistical fits but reports are slightly clearer, with a persistent need to improve stopping criteria, default tolerance settings or messages. 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 more justified.

This paper has also updated some comments about OpenOffice.org Calc contained in Yalta (2008) and Almiron et al. (2010). Of course users have more incentives in suggesting changes in an 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.

8. References

References

Almiron, M. G., B. Lopes, A. L. C. Oliveira, A. C. Medeiros and A. C. Frery (2010), On the numerical accuracy of spreadsheets, Journal of Statistical Software, 34, 4, 1-29.

Bailey, D. H. (1993), A portable high performance multiprecision package. RNR Tech. Rep. 90-022, NASA Ames Research Center, Moffett Field, Calif.

Berger R. L. (2007), Nonstandard operator precedence in Excel, Computational Statistics and Data Analysis, 51, 2788-2791.

Crabtree, C. (2009), How random is random http://www.cliffcrabtree .com/index.php? option=com_content.

Cryer, J. D. (2001), Problems with using Microsoft Excel for statistics, Proceedings of the Joint Statistical Meetings, American Statistical Association, Atlanta.

Fylstra, D., Lasdon, L., Watson, J., and Waren, A. D. (1998), Design and use of the Microsoft Excel Solver, Interfaces, 28, 29-55.

Fylstra, D. and Gupta, S. (2011), Personal communication.

Hargreaves B. R. and T. P. McWilliams (2010), Polynomial trendline function flaws in Microsoft Excel, Computational Statistics and Data Analysis, 54, 1190-1196.

Heiser, D. A. (2009), Microsoft Excel 2000, 2003 and 2007 faults, problems, workarounds and fixes, http://www.daheiser.info/excel/frontpage.html.

Hümmer,T. (2010), International OpenOffice market shares, http://www.webmasterpro.de/portal/news/2010/02/05/ international -openoffice-market-shares.html .

IBM (2010), The risks of using spreadsheets for statistical analysis, IBM Software Business Analytics, Somers, http://public.dhe. ibm.com/common/ssi/ecm/en/imw14297usen/IMW14297USEN.PDF.

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

Knüsel, L. (2003), Computation of statistical distributions - documentation of program ELV, 2nd edition, http://www.stat.uni-muenchen.de/~knuesel.

Knüsel, L. (2005), On the accuracy of statistical distributions in Microsoft Excel 2003, Computational Statistics and Data Analysis, 48, 445-449.

Koenker, R. (2005), Quantile Regression, Cambridge University Press, Cambridge.

Lasdon, L. S., Waren, A. D., Jain and Ratner, M. (1978), Design and testing of a generalized reduced gradient code for nonlinear programming, ACM Transactions of Mathematical Software, 4, 34-50.

L'Ecuyer, P. and R. Simard (2007), TESTU01: A C library for empirical testing of random number generators, ACM Transactions on Mathematical Software 33, 22/1-40.

McCullough B. D. (1998), Assessing the reliability of statistical software: part I, The American Statistician, 52, 358-366.

McCullough B. D. (1999), Assessing the reliability of statistical software: part II, The American Statistician, 53, 149-159.

McCullough B. D. (2000), The accuracy of Mathematica 4 as a statistical package, Computational Statistics, 15, 279-299.

McCullough, B. D. (2004a), Some details of nonlinear estimation, In Micah Altman, Jeff Gill and Michael P. McDonald, editors, Numerical Issues in Statistical Computing for the Social Scientist, Chapter 8, New York: Wiley, pp 199-218.

McCullough, B. D. (2004b), Wilkinson's tests and econometric software, Journal of Economic and Social Measurement, 29, 261-270.

McCullough, B. D. (2004c), Fixing Statistical Errors in Spreadsheet Software: The Cases of Gnumeric and Excel, CSDA Statistical Software Newsletter, www.csdassn.org/software reports.cfm.

McCullough, B. D. (2006), A review of TESTU01, Journal of Applied Econometrics, 21, 677-682.

McCullough, B. D. (2008a), Editorial: Special section on Microsoft Excel 2007, Computational Statistics and Data Analysis, 52, 4568-4569.

McCullough, B. D. (2008b), Microsoft Excel 2007's 'Not The Wichmann-Hill' random number generator, Computational Statistics and Data Analysis, 52, 4587-4593.

McCullough B. D. and D. A. Heiser (2008), On the accuracy of statistical procedures in Microsoft Excel 2007, Computational Statistics and Data Analysis, 52, 4570-4578.

McCullough B. D. and Renfro C. R. (2000), Some numerical aspects of nonlinear estimation, Journal of Economic and Social Measurement, 26, 63-77.

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

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

McCullough B. D. and B. Wilson (2005), On the accuracy of statistical procedures in Microsoft Excel 2003, Computational Statistics and Data Analysis, 49, 1244-1252.

Matsumoto M. and T. Nishimura (1998), Mersenne Twister: a 623-dimensionally equidistributed uniform pseudorandom number generator, ACM Transactions on Modeling and Computer Simulations, 8, 3-30.

Mélard, G. (2011), Utiliser un tableur dans l'enseignement de la statistique. Pourquoi et comment, submitted.

Microsoft Office (2009), 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

Microsoft Office Blog (2010a), Function improvements in Excel 2010, http://blogs.msdn.com/b/excel/archive/2009/09/10/function-improvements-in-excel-2010.aspx.

Microsoft Office Blog (2010b), New and improved solver, http://blogs.office.com/b/microsoft-excel/archive/2009/09/21/new-and-improved-solver.aspx.

Microsoft Office (2010c), About Solver, http://office.microsoft.com/en-us/excel-help/about-solver-HP005198368.aspx?pid=CH010004571033

Nash J. C. (2008), Teaching statistics with Excel 2007 and other spreadsheets, Computational Statistics and Data Analysis, 52, 4602-4606.

Stochastic Global Optimization Methods. Part I: Clustering methods, Mathematical Programming, 39, 27-56.

Stochastic Global Optimization Methods. Part II: Multi level methods, Mathematical Programming, 39, 57-78.

Sawitzki G. (1994), Report on the Numerical Reliability of Data Analysis Systems, Computational Statistics and Data Analysis, 18, 2, 289-301.

Simard, R. (2011). Personal communication.

Simon, G. (2000). https://www.jiscmail.ac.uk/cgi-bin/wa.exe?A2=ind0012& L= ASSUME &F=&S=&X=31BDFD6E4E1E0E4F11&P=4325.

Su, Y.-S. (2008), It's easy to produce chartjunk using Microsoft Excel 2007 but hard to make good graphs Computational Statistics and Data Analysis, 52, 4594-4601.

Ueberhuber, C. W. (1946), Numerical computation: methods, software, and analysis, Springer, Berlin.

Wichmann, B. A. and I. D. Hill (1982). Algorithm AS 183: An efficient and portable pseudo-random number generator, Journal of the Royal Statistical Society Series C Applied Statistics, 31 188-190. Reprinted with correction in Hill I. D. and P. Griffith (Eds) Applied Satistics Algorithms, Ellis Horwood, Chichester, 1985.

Yalta, A. T. (2008), The accuracy of statistical distributions in Microsoft Excel 2007, Computational Statistics and Data Analysis, 52, 4579-4586.

Yalta, A. T. and Yalta A. Y. (2010), Should economists use open source software for doing research? Computational Econometrics 35, 371-394.

Yalta, A. T. and Jenal O. (2009), On the importance of verifying forecasting results. International Journal of Forecasting, 25, 62-73.