Journal of Statistical Software

January 2013, Volume 52, Issue 7. *http://www.jstatsoft.org/*

Spreadsheets in the Cloud – Not Ready Yet

Bruce D. McCullough
Drexel University

A. Talha Yalta
TOBB University of Economics and Technology

Abstract

Cloud computing is a relatively new technology that facilitates collaborative creation and modification of documents over the internet in real time. Here we provide an introductory assessment of the available statistical functions in three leading cloud spreadsheets namely **Google Spreadsheet**, Microsoft **Excel Web App**, and **Zoho Sheet**. Our results show that the developers of cloud-based spreadsheets are not performing basic quality control, resulting in statistical computations that are misleading and erroneous. Moreover, the developers do not provide sufficient information regarding the software and the hardware, which can change at any time without notice. Indeed, rerunning the tests after several months we obtained different and sometimes worsened results.

Keywords: cloud computing, spreadsheet, accuracy, **Google Docs**, **Excel**, **Zoho**, Wilkinson tests.

1. Introduction

Spreadsheets are multipurpose tools commonly used for performing all sorts of computations, including those that involve statistics and data analysis. These programs are based on presenting and manipulating data via an intuitive two-dimensional matrix interface, which is arguably the main reason for their tremendous popularity. A new trend that can potentially make spreadsheet software even more indispensable is the emergence of cloud based alternatives. These are online applications that offer innovative features such as simultaneous multi-user collaboration, real-time data updates from remote sources, as well as improved security and accessibility.

As promising as they may be, an implicit assumption behind the new cloud spreadsheets is that they perform accurately; that is, that users can trust them. In our extensive experience with assessing the accuracy of statistical software in general, and of PC-based spreadsheet software specifically, we have not found such assumptions to be true. Despite their popularity, PC-

based spreadsheets cannot be trusted to perform statistical calculations correctly (Almiron, Lopes, Oliveira, Medeiros, and Frery 2010; Knüsel 2002; McCullough 2008b,a; McCullough and Heiser 2008; McCullough and Wilson 1999, 2002, 2005; Yalta 2007, 2008; Yalta and Jenal 2009). A notable exception discussed by McCullough (2004a) is the open-source spreadsheet program **Gnumeric** (The **Gnumeric** Team 2012).

Spreadsheet use is moving to the cloud; the cloud office suites offered by Google and Microsoft claim 25 and 20 million users, respectively (Press Release Point 2010). As spreadsheets move to the cloud, it is not unreasonable to expect that cloud users will perform myriad statistical operations using cloud spreadsheets. Indeed, Yalta (2008) performed a cursory examination of **Google Spreadsheet** (Google Inc. 2012), reported "gross errors," and recommended a detailed evaluation to help researchers and practitioners make the decision whether to move from the PC to the cloud. This raises the obvious question, "Can these cloud spreadsheets be trusted to perform calculations correctly?" The answer may very well be "no."

Here we present results from thorough testing of three leading cloud spreadsheets namely **Google Spreadsheet**, Microsoft **Excel Web App** (Microsoft Corporation 2012), and **Zoho Sheet** (**Zoho** Corporation 2012). Our methodology is based on applying Wilkinson's *Statistics Quiz* (1985) for assessing statistical functionality in general, and using the Knüsel (1989) approach for auditing statistical distributions in particular. The results show that these cloud spreadsheets cannot currently be trusted to perform statistical operations with accuracy.

It is important to understand that these are not "gotcha" tests designed simply to trip up software developers. These are carefully designed tests designed to inform the user about the quality of the software. As McCullough (2004b) wrote about Wilkinson's Tests: "The flaws they are designed to expose have well-known solutions. That is, these are tests which any package *could* pass. If a software package fails a particular test, there exists a known method of obtaining the correct answer." The same is true of the tests of the accuracy of the statistical distributions.

2. "Statistics Quiz"

Statistics Quiz (Wilkinson 1985) is a well-known collection of test problems that are designed to expose flaws in statistical software. It comprises 20 tests (some with sub-parts) in six areas: (I) reading an ASCII file; (II) handling real numbers; (III) handling missing data; (IV) regression; (V) analysis of variance; and (VI) operating on a database. The tests are based on a small and effective data set called "Nasty." As can be seen in Table 1, all of the numbers in `nasty.dat` are within the range of representable numbers for 32-bit double precision. The resulting tests aim to provide a standardized approach to expose common flaws for which there are well-known and easily implemented solutions. Thanks to its efficiency, *Statistics Quiz* has been employed by various authors such as Sawitzki (1994); Bankhofer and Hilbert (1997); McCullough (2004b); Choi and Kiefer (2005); Yalta and Yalta (2009). It has been important in the detection and correction of dozens of errors in statistical packages as well as spreadsheet software.

We applied *Statistics Quiz* to the following cloud spreadsheets: **Google Spreadsheet**, **Excel Web App**, and **Zoho Sheet**. We skipped Test I (A, B), Test II-E, Test V (A, B, C), and Test VI (A, B), which are not relevant and/or not applicable with spreadsheet software. The tests were performed between May 2–6 2011, and using an Intel Core i7 2.80GHz desktop computer

LABEL$	X	ZERO	MISS	BIG	LITTLE	HUGE	TINY	ROUND
ONE	1	0	.	99999991	0.99999991	1.0E12	1.0E−12	0.5
TWO	2	0	.	99999992	0.99999992	2.0E12	2.0E−12	1.5
THREE	3	0	.	99999993	0.99999993	3.0E12	3.0E−12	2.5
FOUR	4	0	.	99999994	0.99999994	4.0E12	4.0E−12	3.5
FIVE	5	0	.	99999995	0.99999995	5.0E12	5.0E−12	4.5
SIX	6	0	.	99999996	0.99999996	6.0E12	6.0E−12	5.5
SEVEN	7	0	.	99999997	0.99999997	7.0E12	7.0E−12	6.5
EIGHT	8	0	.	99999998	0.99999998	8.0E12	8.0E−12	7.5
NINE	9	0	.	99999999	0.99999999	9.0E12	9.0E−12	8.5

Table 1: The data set `nasty.dat`.

running a fully updated Windows 7 Home Premium operating system and the Firefox 4 web browser.

The results for the 12 applied tests are as follows.

2.1. Test II-A

In this test, `ROUND` is printed with only one digit and the expected correct answer is the numbers from 1 to 9. Depending on the language compiler used, this operation is sometimes performed automatically using the "round-to-even" method, resulting in unacceptable output such as, for example, `R(1.5) = R(3.5) = 2`. In addition, the following expressions are checked:

```
Y1 = INT(2.6*7 -0.2)
Y2 = 2-INT(EXP(LOG(SQR(2)*SQR(2))))
Y3 = INT(3-EXP(LOG(SQR(2)*SQR(2))))
```

where the correct answers are 18, 0, and 1 respectively.

Conclusion: All three packages pass this test.

2.2. Test II-B

This test involves plotting `HUGE` against `TINY`, `BIG` against `LITTLE`, and `X` against `ZERO`. From looking at the data, it is obvious that the first two plots should be a straight line with a 45 degree angle, and the third plot should be a vertical line. **Zoho Sheet** passes this test with all the graphs looking good except the disappearing tick-marks in the horizontal axis in `BIG` against `LITTLE`. The test is not applicable to **Excel Web App**, which does not support creating and editing graphs online. Finally, **Google Spreadsheet** fails this task due to plotting `BIG` against `LITTLE` as a horizontal line. The other 2 graphs are passable. Figure 1 presents the `BIG` against `LITTLE` plots produced by **Zoho Sheet** and **Google Spreadsheet** respectively.

Conclusion: **Zoho Sheet** passes, **Google Spreadsheet** fails.

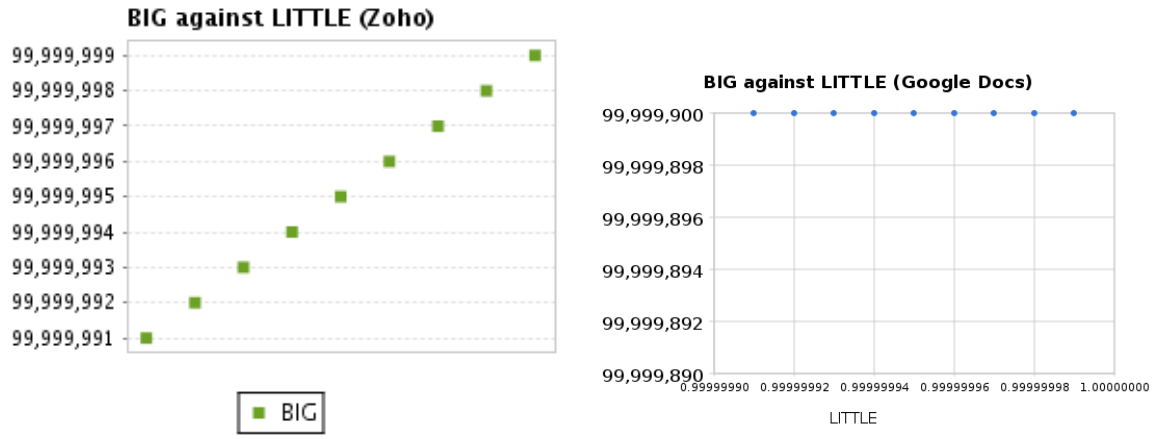

Figure 1: Wilkinson's tests II-B results for **Zoho Sheet** and **Google Spreadsheet**.

2.3. Test II-C

Here, the mean and the standard deviation are computed for all of the series. The expected result is the fifth value as the mean for all variables. Also, the standard deviations should be 2.738612788 times 10 to a power for all variables except `ZERO` and `MISS`. The correct answer is 0 for `ZERO`, and either "undefined" (`#DIV/0!`) or "missing" (`#N/A`) for `MISS`.

Excel Web App and **Zoho Sheet** return the correct values. **Google Spreadsheet** fails with standard deviations accurate to only 1 significant digit. These results are presented in Table 2. Traditional 32-bit double precision is more than sufficient to produce the correct answer, and accurate methods for computing this quantity have long been known (Chan, Golub, and Leveque 1983). We have no idea how Google is performing this calculation.

Conclusion: **Zoho Sheet** and **Excel Web App** pass, **Google Spreadsheet** fails.

2.4. Test II-D

Calculate the correlation matrix of the variables. This should be unity for all correlations except those involving `ZERO` and `MISS`. Mathematically (and computationally), correlations should be between −1 and 1. Yet, **Google Spreadsheet** manages to compute the correlation between `X` and `BIG` as 1.19 and the correlation between `BIG` and itself as 1.41. Again, it is

Variable	Correct	**Google Spreadsheet**
`X`	2.73861E+00	2.74
`ZERO`	na	0
`MISS`	na	#N/A
`BIG`	2.73861E+00	2.**3094010767585**
`LITTLE`	2.73861E−08	0.0000000**8097899**
`HUGE`	2.73861E+12	2,738,612,787,525.83
`TINY`	2.73861E−12	0.000000000002739
`ROUND`	2.73861E+00	2.73861278752583

Table 2: Standard deviation calculations for **Google Spreadsheet**; inaccurate digits in bold.

obvious that **Google Spreadsheet** is using a bad algorithm or has not properly programmed a good algorithm.

Conclusion: **Zoho Sheet** and **Excel Web App** pass, **Google Spreadsheet** fails.

2.5. Test II-F

Regress `BIG` on `X`. The intercept should be 99999990 and the slope should be unity. All of the packages reproduce this result. Curiously, though, **Zoho Sheet** manages to compute sum of squared residuals $= -13$, $R^2 = 1.28$, and F ratio $= -32.31$. These should be 0, 1, and either "undefined" (`#DIV/0!`) or "missing" (`#N/A`) respectively. We would be interested to know what formula **Zoho Sheet** is using to compute these statistics.

Conclusion: **Zoho Sheet** fails, **Excel Web App** and **Google Spreadsheet** pass.

2.6. Test III-A

This test involves the transformation

```
IF MISS = 3 THEN TEST = 1 ELSE TEST = 2
```

where the acceptable answers are either 2 or "missing" (`#N/A`). We performed this with the command `=IF((VALUE=3);1;2)` using `MISS` as the reference for `VALUE`.

Conclusion: All three packages pass this test.

2.7. Test III-B

This test involves the computation

```
IF MISS = <missing> THEN MISS = MISS + 1
```

which has the correct answer `<missing>` since one cannot add 1 to something that is missing. We performed this computation using the command `=IF((ISNA(VALUE)=TRUE);(D3+1);)` using `MISS` as the reference for `VALUE`.

Conclusion: All three packages pass this test.

2.8. Test IV-A

In this test, `X` is regressed on a constant plus the variables $\texttt{X}^2$ through $\texttt{X}^9$ in view of the fact that an unstable regression function often cannot handle all of these polynomials. Because the objective of this pragmatic test is the overall regression, the primary focus is on checking whether all of the standard errors are 0 and R^2 is unity.

It is important to note that, in spreadsheets, this type of problem usually is handled by creating an array formula such as `LINEST()` in **Excel**. **Google Spreadsheet** permits this operation, but **Excel Web App** does not allow the creation of array formulae. However, if an **Excel** spreadsheet with this operation is uploaded, **Excel Web App** computes the answer. **Zoho Sheet** does not even list `LINEST` among the available functions, yet computes results if a spreadsheet with `LINEST` is uploaded.

Google Spreadsheet fails to run this regression and returns `#REF!` as a misleading error message. Adding the variables one at a time, we understand that the program cannot handle

Package	c	b_1	b_2
Google Spreadsheet	−125829118	0	1.26
Excel Web App	−99999990	0	1
Zoho Sheet	−96199094.6	8374388.9	0.07825

Table 3: Results for Test IV-C.

`X`[9]. For **Excel Web App** and **Zoho Sheet**, setting up the test first offline and then importing into the programs reveals that **Excel Web App** passes this test. **Zoho Sheet**, however, fails to compute the standard errors, returning the error `#VALUE!` instead.

Conclusion: **Excel Web App** passes, **Zoho Sheet** and **Google Spreadsheet** fail.

2.9. Test IV-B

Here, `X` is regressed on `X` to see whether the program runs without complaining about this perfectly valid operation, and returns the obvious solution `X` $= 0 + 1$`X` with $R^2 = 1$ and $F_{1,7} =$ undefined (`#VALUE!`).

In this case, **Google Spreadsheet** and **Zoho Sheet** returns the correct values for each statistic while **Excel Web App** fails by returning $F_{1,7}$ =1.58E+032. Unlike other programs, **Excel Web App** also does not round to zero and returns the values 8.88E−016, 7.96E−017, 4.48E−016, 2.66E−030 for the constant, standard errors, and the sum of squared residuals respectively.

Conclusion: **Zoho Sheet** passes, **Excel Web App** fails, **Google Spreadsheet** passes.

2.10. Test IV-C

Regress `X` on `BIG` and `LITTLE`. Since all the variables are perfectly collinear, the design matrix is singular and an infinite number of solutions exist. Accordingly, the program should produce an error message, preferably one that diagnoses and warns of the singularity. Mathematically, the formula is `X` $= c + b_1$ `BIG` $+ b_2$ `LITTLE`. The "solutions" returned by the three packages are presented in Table 3: three packages, three answers, all of them wrong. On a historical note, we mention that not only did **Excel** 4.0 released in 1992 fail this test (Sawitzki 1994), but that **Excel Web App** released in 2009 still fails it, and so does **Excel** 2010.

Conclusion: All three packages fail this test.

2.11. Test IV-D

In this test, `ZERO` is regressed on a constant and `X`. The acceptable behavior is either the program notifies that `ZERO` has no variance, or reports regression output showing `ZERO` $= 0+0$`X` with total sum of squares equaling zero.

Again, for **Web App** and **Zoho Sheet**, we run this test by first setting it up offline on a PC-based spreadsheet, and then uploading it to the cloud. None of the packages warns about the dependent variable, however, the regression output of the three programs shows the expected values.

Conclusion: All three packages pass this test.

3. Statistical distributions

Depending on the field or industry, spreadsheet users may make extensive use of statistical distributions. That is why spreadsheets offer so many distributions, not just a few. On the other hand, these are important decision making tools and it is important that they be accurate or not offered at all. In today's computing standards, accuracy means that any program offering statistical distributions be able to reliably compute tail probabilities as small as 10^{-100} with at least 6 significant digits. This is perfectly possible on commodity hardware, and there are known and open source algorithms that can provide such precision (Yalta 2008; Bangalore, Wang, and Allison 2009).

The de-facto benchmarking tool for assessing the accuracy of statistical distributions is the program **ELV** (Knüsel 1989). Like *Statistics Quiz*, **ELV** has been employed in the past by numerous studies such as McCullough and Wilson (1999, 2002, 2005); Knüsel (1995, 1998, 2002, 2005); Bustos and Frery (2006); Yalta (2008). We use the **ELV** program to assess the reliability of a number of statistical distributions for which a dedicated function is available in **Zoho Sheet**, **Excel Web App**, and **Google Spreadsheet**. The function names are the same in the three spreadsheets where available, however, the algorithms are apparently different. Also, for the binomial and χ^2 distributions, **Zoho Sheet** offers alternative, more flexible functions named `B` and `CHISQDIST` respectively. We do not assess these two functions and consider only the **Excel** compatible versions that are automatically used in all imported documents.

As can be seen in Table 4, for lower tails of the standard normal distribution, all three packages can return seriously misleading results with 0 digits of accuracy for not so small probabilities. **Google Spreadsheet** can even give negative p values.

Google Spreadsheet does not offer the Student's t, χ^2, or F distributions. For t and F, **Zoho Sheet** can return 0 for not so small probabilities (on the order of 1E−5). It can also produce misleading results with only 2 digits of accuracy. **Excel Web App** appears to be accurate for the three distributions.

The discrete probability distributions binomial, hypergeometric, and Poisson are available in **Google Spreadsheet**, however, none of these is reliable. For Poisson, **Google Spreadsheet** can

z	**ELV** (exact)	**Zoho Sheet**	**Excel Web App**	**Google Spreadsheet**
0	0.5	exact	exact	exact
−1	1.58655E−01	exact	exact	exact
−2	2.27501E−02	exact	exact	exact
−3	1.34990E−03	exact	exact	exact
−4	3.16712E−05	exact	exact	3.16698E−05
−5	2.86652E−07	2.86672E−07	exact	−2.36199E−07
−6	9.86588E−10	9.86600E−10	exact	9.93806E−10
−7	1.27981E−12	1.27982E−12	exact	7.24218E−08
−8	6.22096E−16	6.22088E−16	exact	0
−8.5	9.47953E−18	2.23346E−17	2.23239E−17	0
−9	1.12859E−19	1.08420E−19	exact	0
−10	7.61985E−24	0	exact	0

Table 4: Results for the standard normal distribution.

p	**ELV** (exact)	**Zoho Sheet**	**Excel Web App**	**Excel** 2010	**Excel** 2007
1E−1	3.98635E+01	exact	exact	exact	exact
1E−2	4.05218E+03	exact	exact	exact	exact
1E−3	4.05284E+05	exact	exact	exact	exact
1E−4	4.05285E+07	exact	exact	exact	exact
1E−5	4.05285E+09	99999999	exact	exact	1000000000
1E−6	4.05285E+11	99999999	exact	4.05292E+11	1000000000
1E−7	4.05285E+13	99999999	exact	4.05730E+13	1000000000
1E−8	4.05285E+15	99999999	exact	4.50360E+15	1000000000
1E−9	4.05285E+17	99999999	exact	#NUM!	1000000000

Table 5: Results for inverse $F(1,1)$ with probability $= p$.

return misleading results for not so small probabilities. It can even give ∞ as p values. It also cannot calculate when the trial number $n > 136$ and the finite population size $N > 175$ for the binomial and the hypergeometric distributions respectively. **Zoho Sheet** and **Excel Web App** appear to be accurate for these three distributions.

Google Spreadsheet does not offer the computation of any inverse distribution function except for the normal distribution, which seems to be accurate. **Zoho Sheet** offers an inverse function corresponding to its statistical distributions, however, these are all unstable and can erroneously return results such as 9999.99995, 99999999, and 0.00000001 for different distributions. For **Excel Web App**, the quantiles of the various distributions seem to be accurate, however, it is also noticeable that the computations can be different from those reported offline by **Excel** 2010.

The above results can best be illustrated with the inverse F distribution. As Table 5 shows, **Zoho Sheet** and **Excel** 2007 both have difficulties in the same place, starting at 1E−5. **Excel** 2010 does only a bit better than **Excel** 2007. What is surprising is that the answers one gets from a Microsoft spreadsheet depend on whether the spreadsheet is opened in **Excel** 2010 or **Excel Web App**. This does not bode well for the idea of portable spreadsheets, even within the Microsoft family. How distressing it would be for a user to find that his results depend on whether his software is PC-based or in the cloud.

4. The issue of accountability

An important issue with cloud based software is in the department of accountability. In a setting where the web browser acts merely as a terminal window to display results computed somewhere in the cloud, the user often has no information on computational details or whether the results can be replicated in the future. Moreover, the user also has no control over the software and the hardware, which is subject to change by the provider at any time and without warning. For obtaining help and support, there are various online resources and services such as official blogs and help forums, however, these services are informal in nature and the information they provide is often dispersed and lacking in terms of technical details.[1]

[1] **Excel Web App** has an official forum located at http://answers.microsoft.com/en-us/office/forum/webapps-excel, as well as an official blog at http://blogs.office.com/b/officewebapps/. **Zoho Sheet** incorporates a "Feedback" button on the top menu. In addition, there is a Zoho Forum, an official Zoho Blog

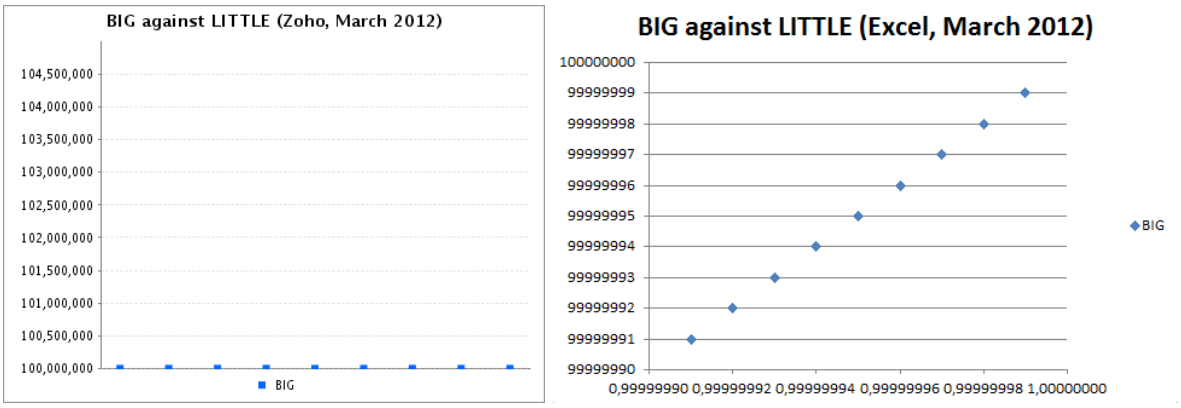

Figure 2: March 2012 Test II-B results for **Zoho Sheet** and **Excel Web App**.

After an extensive scrutinization of the various available online resources, we came to understand that the three cloud spreadsheets analyzed in this study come neither with a version number nor any information that we could find regarding the machine specifications and the operating system used. We also realized that, over the weeks after our initial testing, major revisions have taken place in all of the three programs and there is no method of replicating any earlier result. In order to assess the practical consequences of the current state of this way of supplying tools for data analysis, we reapplied the tests on March 1–2, 2012. Here is what we found:

Zoho Sheet now lists `LINEST()` among the available functions. From the change log, we understand that this feature has been implemented in August 2011. On the other hand, as can be seen in Figure 2, **Zoho Sheet** can no longer correctly produce the `BIG` against `LITTLE` plot in Test II-B. We do not know when this has happened. The change log announces chart "enhancements" in both August 2011 and January 2012, however the explanations supplied are regarding cosmetic changes such as the colors and the fonts used.

Excel Web App can now create/edit graphs online, and it passes Test II-B by correctly producing all of the three plots including the relatively more difficult `BIG` against `LITTLE` plot shown in Figure 2. It also now gets the correct probability for $z = -8.5$ for the standard normal distribution. On the other hand, **Excel Web App** now fails Test II-F by computing 4.48E−16 and 7.96E−17 for the standard error of $\hat{\beta}_0$ and $\hat{\beta}_1$, both of which should be 0. Also the sum of squared residuals is reported as 2.66E−30, not 0. According to the official blog, the revisions leading to these discrepancies have likely took place in a general update in September 2011, which announces "a lot of behind-the-scenes improvements" in **Office Web Apps**.

Apparently, since our initial tests in May 2011, **Google Docs** has also received various updates and there is now a "new version" of **Google Spreadsheet**. We were unable to find out what version this is, however, according to the support website "An easy way to tell if you're using the new version is if there's a ruler above the editing space. (Just be sure the ruler is enabled from View > Show ruler)" From the official **Google Docs** blog we infer that the switch to the

and a change log; which are located at https://forums.zoho.com/, http://blogs.zoho.com, and https://sheet.zoho.com/features respectively. **Google Spreadsheet** offers a help forum, an official **Google Docs** Forum, and a support website; which can be found at http://groups.google.com/a/googleproductforums.com/forum/#!forum/docs, http://googledocs.blogspot.com/, and http://support.google.com/docs .

Tests	**Excel Web App**	**Zoho Sheet**	**Google Spreadsheet**
II-A	pass	pass	pass
II-B	na (pass)	pass (**fail**)	**fail**
II-C	pass	pass	**fail**
II-D	pass	pass	**fail**
II-F	pass (**fail**)	**fail**	pass
III-A	pass	pass	pass
III-B	pass	pass	pass
IV-A	pass	**fail**	**fail**
IV-B	**fail**	pass	pass
IV-C	**fail**	**fail**	**fail**
IV-D	pass	pass	pass

Distributions	**Excel Web App**	**Zoho Sheet**	**Google Spreadsheet**
Standard Normal	**fail** (pass)	**fail**	**fail**
t	pass	**fail**	na
χ^2	pass	pass	na
F	pass	**fail**	na
Binomial	pass	pass	**fail**
Hypergeometric	pass	pass	**fail**
Poisson	pass	pass	**fail**
Inverse Normal	pass	**fail**	pass
Inverse t	pass	**fail**	na
Inverse χ^2	pass	**fail**	na
Inverse F	pass	**fail**	na

Table 6: Results summary for Wilkinson tests and statistical distributions. The changes after March 2012 tests are shown in parentheses.

"new version" has taken place in October 2011.

5. Conclusion

The relatively new cloud spreadsheets bring important innovations such as real-time data updates and the ability to simultaneously view and make changes to documents by multiple users. However, these programs are first and foremost designed for ease of use, and the fact that scientists constitute only a small fraction of their target audience creates important concerns in the department of accuracy and replicability. Offering statistical functions in a computer program is a serious matter and it is necessary that such functions are extremely accurate and reliable, or not offered at all. After all, these are important tools used for inference and decision making. Indeed, Panko and Ordway (2005) provide ample evidence that the vast majority of financial and business decisions, including some that have caused much havoc for world markets, are in some way made using spreadsheets. Poor statistical tools offered for the sake of compatibility or feature parity with other spreadsheets will do nothing but increase computational garbage and information pollution.

Wilkinson's tests (1985) and Knüsel's approach (1989) are entry level test procedures commonly used to assess statistical functionality in general and statistical distributions in particular. We have employed these two methods to perform an all-around assessment of three leading cloud-based spreadsheets namely **Google Spreadsheet**, Microsoft **Excel Web App**, and **Zoho Sheet**. The summary of the results are presented in Table 6. As can be seen, there are errors and/or deficiencies in all of the three programs and in such areas as univariate and multivariate summary statistics, visualization, regression, and statistical distributions. It is particularly striking that **Zoho Sheet** fails about half of the tests while **Google Spreadsheet** has more failures than passes. Also, despite being under development for about 25 years, and the numerious earlier reports pointing out the accuracy issues, **Excel** still has failures in these simple tests, both offline and in the cloud. Moreover, the output can differ depending on whether one uses the PC-based **Excel** 2010 or the cloud based **Excel Web App**. Our findings are also consistent with Keeling and Pavur (2011), who perform a detailed testing (including a subset of Wilkinson tests) of **Google Spreadsheet** along with five other PC-based spreadsheets and report that "it is not recommended for statistical analysis beyond simple averages."

An important concern regarding the cloud spreadsheets is that the user has limited or no information on the software and the hardware, which are subject to change by the provider at any time and without warning. Indeed, rerunning the tests after several months, we obtained the different results shown in parentheses in Table 6. These changes indicate that, in the absence of version numbers as well as sufficient information on the computing environment, these programs can not be used for anything that requires verification, such as most business functions.

Based on the findings, it is our understanding that the developers of cloud-based spreadsheets have not been performing basic quality control and providing sufficient documentation, and that the results of their statistical calculations cannot be trusted. Whether the user can expect the developers to correct these problems is a legitimate concern. Zoho and Google have no track record when it comes to fixing errors in spreadsheets, but the errors they have made demonstrate a complete lack of knowledge concerning the literature on the accuracy of statistical software, in particular, spreadsheets. Microsoft has a long track record of failing to fix errors in the PC version of **Excel**, including **Excel** 2010 (Knüsel 2011; Mélard 2011). Whether Microsoft can fix errors in its cloud spreadsheet without having fixed errors in its PC spreadsheet we are doubtful. We leave it to the reader to decide whether other calculations performed by cloud-based spreadsheets can be trusted.

Acknowledgments

We would like to thank the anonymous reviewers for extremely useful comments.

References

Almiron MG, Lopes B, Oliveira ALC, Medeiros AC, Frery AC (2010). "On the Numerical Accuracy of Spreadsheets." *Journal of Statistical Software*, **34**(4), 1–29. URL http://www.jstatsoft.org/v34/i04/.

Bangalore SS, Wang J, Allison DB (2009). "How Accurate Are the Extremely Small p-

Values Used in Genomic Research: An Evaluation of Numerical Libraries." *Computational Statistics & Data Analysis*, **53**, 2446–2452.

Bankhofer U, Hilbert A (1997). "Statistical Software Packages for Windows – A Market Survey." *Statistical Papers*, **38**, 377–471.

Bustos OH, Frery AC (2006). "Statistical Functions and Procedures in **IDL** 5.6 and 6.0." *Computational Statistics & Data Analysis*, **50**, 301–310.

Chan TF, Golub GH, Leveque RJ (1983). "Algorithms for Computing the Sample Variance: Analysis and Recommendations." *The American Statistician*, **37**, 242–247.

Choi HS, Kiefer NM (2005). "Software Evaluation: **EasyReg** International." *International Journal of Forecasting*, **21**, 609–616.

Google Inc (2012). ***Google Spreadsheet***. Mountain View. URL https://spreadsheets.google.com/.

Keeling KB, Pavur RJ (2011). "Statistical Accuracy of Spreadsheet Software." *The American Statistician*, **65**, 265–273.

Knüsel L (1989). *Computergestützte Berechnung Statistischer Verteilungen*. Universität München Verlag, München.

Knüsel L (1995). "On the Accuracy of the Statistical Distributions in GAUSS." *Computational Statistics & Data Analysis*, **20**, 699–702.

Knüsel L (1998). "On the Accuracy of Statistical Distributions in Microsoft **Excel** 97." *Computational Statistics & Data Analysis*, **26**, 375–377.

Knüsel L (2002). "On the Reliability of Microsoft **Excel** XP for Statistical Purposes." *Computational Statistics & Data Analysis*, **39**, 109–110.

Knüsel L (2005). "On the Accuracy of Statistical Distributions in Microsoft **Excel** 2003." *Computational Statistics & Data Analysis*, **48**, 445–449.

Knüsel L (2011). "On the Accuracy of Statistical Distributions in Microsoft **Excel** 2010." CSDA Statistical Software Newsletter, URL http://www.csdassn.org/software_reports/Excel2011.pdf.

McCullough BD (2004a). "Fixing Statistical Errors in Spreadsheet Software: The Cases of **Gnumeric** and **Excel**." CSDA Statistical Software Newsletter, URL http://www.csdassn.org/software_reports/gnumeric.pdf.

McCullough BD (2004b). "Wilkinson's Tests and Econometric Software." *Journal of Economic and Social Measurement*, **29**, 261–270.

McCullough BD (2008a). "Microsoft **Excel**'s 'Not the Wichmann–Hill' Random Number Generators." *Computational Statistics & Data Analysis*, **52**, 4587–4593.

McCullough BD (2008b). "Special Section on Microsoft **Excel** 2007." *Computational Statistics & Data Analysis*, **52**, 4568–4569.

McCullough BD, Heiser DA (2008). "On the Accuracy of Statistical Procedures in Microsoft **Excel** 2007." *Computational Statistics & Data Analysis*, **52**, 4570–4578.

McCullough BD, Wilson B (1999). "On the Accuracy of Statistical Procedures in Microsoft **EXCEL** 97." *Computational Statistics & Data Analysis*, **31**, 27–37.

McCullough BD, Wilson B (2002). "On the Accuracy of Statistical Procedures in Microsoft **Excel** 2000 and **Excel** XP." *Computational Statistics & Data Analysis*, **40**, 713–721.

McCullough BD, Wilson B (2005). "On the Accuracy of Statistical Procedures in Microsoft **Excel** 2003." *Computational Statistics & Data Analysis*, **49**, 1244–1252.

Mélard G (2011). "On the Accuracy of Statistical Procedures in Microsoft **Excel** 2010." Working paper, URL http://homepages.ulb.ac.be/~gmelard/rech/gmelard_csda23.pdf.

Microsoft Corporation (2012). ***Excel Web App***. Redmond. URL http://office.microsoft.com/en-us/web-apps/.

Panko R, Ordway N (2005). "Sarbanes-Oxley: What about All the Spreadsheets?" In D Ward (ed.), *Managing Spreadsheets in the Light of Sarbanes-Oxley*, pp. 15–59. European Spreadsheets Risks Interest Group, London.

Press Release Point (2010). "Microsoft **Office Web Apps** Has 20 Million Users." The Seattle Times, accessed 2010-10-06.

Sawitzki G (1994). "Report on the Numerical Reliability of Data Analysis Systems." *Computational Statistics & Data Analysis*, **18**, 289–301.

The **Gnumeric** Team (2012). ***Gnumeric*** *1.11.2*. URL http://projects.gnome.org/gnumeric/.

Wilkinson L (1985). *Statistics Quiz: Problems Which Reveal Deficiencies in Statistical Programs*. SYSTAT, Evanston, IL.

Yalta AT (2007). "The Numerical Reliability of GAUSS 8.0." *The American Statistician*, **61**, 262–268.

Yalta AT (2008). "The Accuracy of Statistical Distributions in Microsoft **Excel** 2007." *Computational Statistics & Data Analysis*, **52**, 4579–4586.

Yalta AT, Jenal O (2009). "On the Importance of Verifying Forecasting Results." *International Journal of Forecasting*, **25**, 62–73.

Yalta AT, Yalta AY (2009). *Wilkinson Tests and* ***gretl***, chapter 16, pp. 243–251. EHUCHAPS. Universidad del Pais Vasco, Departamento de Economia Aplicada III. URL http://ideas.repec.org/h/ehu/ehucha/01-16.html.

Zoho Corporation (2012). ***Zoho Sheet***. Pleasanton, CA. URL https://sheet.zoho.com/.

Affiliation:

B. D. McCullough
Drexel University
Department of Decision Sciences
Philadelphia, PA 19104-2875, United States of America
E-mail: bdm25@drexel.edu
URL: http://www.pages.drexel.edu/~bdm25/

A. Talha Yalta
TOBB University of Economics and Technology
Department of Economics
Sogutozu Caddesi No:43
06560, Ankara, Turkey
E-mail: yalta@etu.edu.tr
URL: http://yalta.etu.edu.tr/index-en.html

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume 52, Issue 7 *Submitted:* 2011-09-20
January 2013 *Accepted:* 2012-10-09