The Numerical Accuracy of Statistical Calculations in Excel 2000 and Minitab Version 13

Terry E. Dielman, Texas Christian University

Certain statistical procedures available in Excel 2000 and Minitab Version 13 are assessed for their numerical reliability. Both Excel and Minitab are widely used in teaching as well as in industry for statistical analysis, so numerical accuracy is important. Minitab is very accurate in the three areas tested: computing univariate statistics, regression, and anova. Excel has shortcomings of which users should be aware. Comparisons are also made to results from previous research on SAS and SPSS. Minitab compares favorably to both these statistical packages.

INTRODUCTION

Recently McCullough (1998) proposed a methodology for assessing the numerical reliability of statistical software using the Statistical Reference Datasets (StRD, available at http://www.nist.gov/itl/div898/strd). These datasets provide a collection of accuracy benchmarks that have been compiled by the National Institute of Standards and Technology (NIST). The data sets are provided along with values for certain statistics certified to a number of decimal places. The certified values can be used to assess the ability of statistical packages to provide accurate computation of these statistics.

In this paper, the procedures available in Excel 2000 and Minitab Version 13 are assessed for their numerical reliability. Since Excel is being used more and more often for educational purposes and in the business world for statistical computations, this type of assessment is important. At the least, recognition of any shortcomings in the numerical reliability should be noted. McCullough and Wilson (1999) examined the accuracy of the statistical procedures in Excel 97 and found them to be lacking. They recommend against the use of Excel 97 for analysis of data. Knüsel (1998) also examined the accuracy of statistical distributions in Excel 97 and found that tail area computations were often inaccurate. This paper extends the results of McCullough and Wilson (1999) to Excel 2000. The procedures in Minitab Version 13 are evaluated against the same benchmarks. Minitab has not been previously evaluated to my knowledge. Minitab is a widely used package for teaching and in industry, so numerical accuracy is important.

THE StRD DATA SETS

The StRD data sets created by NIST have been developed according to level of difficulty. Three levels of difficulty are available, denoted Low (L), Average (A) and High (H).

Univariate Statistics

There are nine data sets in the univariate suite. Six are classified as L: Michelso, PiDigits, Lew, Lottery, Mavro and NumAcc1; two as A: NumAcc2 and NumAcc3; and one as H: NumAcc4. For each data set, certified values are provided for the mean, standard deviation and the first-order autocorrelation coefficient to 15 digits. In this paper, only the mean and standard deviation will be examined.

Anova Tests

There are eleven data sets in the anova suite. Four are classified as L:SiRstv, SmLs01, SmLs02, and SmLs03; four as A:AtmWtAg, SmLs04, SmLs05, and SmLs06; and three as H: SmLs07, SmLs08, and SmLs09. Each data set is a one-way analysis of variance problem. For each data set, certified values are provided for the Abetween treatment@ degrees of freedom, Awithin treatment@ degrees of freedom, sums of squares, mean squares, the F-statistic, the R^2, and the residual standard deviation to 15 digits. Since most of the certified values are used in calculating the F-statistic, only this statistic will be examined.

Linear Regression

There are eleven data sets in the linear regression suite. Two are classified as L: Norris and Pontius; two as A: NoInt1 and NoInt2; and seven as H: Filip, Longley, Wampler1, Wampler2, Wampler3, Wampler4 and Wampler5. For each data set, certified values are provided for the coefficient estimates, standard errors of coefficients, the residual standard deviation, R^2, and the usual analysis of variance for linear regression table to 15 digits. The results shown are for the coefficient estimates and the standard errors of the coefficients.

RESULTS

The accuracy measure used throughout this paper is the negative base 10 logarithm of the relative error (LRE) defined as

$$LRE = -\log_{10}\left(\frac{|q-c|}{|c|}\right) \quad (1)$$

where q is the estimated value and c is the correct value. When q = c, the LRE is undefined, in which case McCullough (1998) suggests setting it equal to the number of digits in c. The LRE will be approximately equal to the number of correct significant digits in the estimate when q is close to c. However, using the LRE avoids certain problems encountered with counting of correct significant digits. The reader is referred to McCullough (1998, p. 360-361) for more detail on the LRE. Values considered acceptable for an LRE vary. McCullough suggests that low-difficulty linear procedures should have an LRE of at least nine, with decreasing LREs being acceptable for more difficult data sets.

Mean

The results for computing the sample mean are shown in Table 1. The certified value is shown for each data set as well as the computed values for Excel and Minitab. This table allows the reader to see the number of correct significant digits in the estimates. The associated LREs are in Table 2. Both Excel and Minitab are accurate in finding the value of the sample mean. The variation in the LREs reflects minor differences with all LREs either 14 or 15. For Excel, these results are the same as in McCullough and Wilson (1999) for Excel 97. In addition to the Excel 2000 and Minitab Version 13 values, the results given by McCullough (1999) for SAS and SPSS are shown.

TABLE 1
Computing the Sample Mean

Data Set	Certified Value	Excel	Minitab
Michelso	299.8524	299.8524	299.8523999999993
Pi Digits	4.5348	4.5348	4.5348.
Lew	-177.435	-177.435	-177.435
Lottery	518.95871559633	518.95871559633	518.95871559633
Mavro	2.001856	2.001856	2.001856
NumAcc1	10000002	10000002	10000002
NumAcc2	1.2	1.199999999999999	1.2
NumAcc3	1000000.2	1000000.2	1000000.1999999994
NumAcc4	10000000.2	10000000.2000001	10000000.2000001

TABLE 2
LREs for Computing the Sample Mean

Data Set	Excel	Minitab	SAS	SPSS
Michelso:	15	14.5	15	15
Pi Digits	15	15	15	14.7
Lew	15	15	15	15
Lottery	15	15	15	15
Mavro	15	15	15	15
NumAcc1	15	15	15	15
NumAcc2	14.0	15	14.0	15
NumAcc3	15	14.0	15	15
NumAcc4	14.0	14.0	14.0	15

Standard Deviation

For computation of the standard deviation, the story is different. The certified values and computed values for both Excel and Minitab are shown in Table 3. The associated LREs are in Table 4. This table also contains the LREs for SAS and SPSS as computed by McCullough (1999). Note that the Excel results are essentially the same as in McCullough and Wilson (1999). Minitab has LREs of 14 and 15 for all but two of the data sets. The average difficulty data set NumAcc3 has an LRE of 9 and the only data set in the H category has an LRE of 8. Excel has LRE=s of 15 on four of the lower difficulty level data sets along with LREs of 8 and 9 on the other two. On the average and high difficulty data sets, Excel's algorithm for standard deviation shows its weaknesses with LREs of 12, 1, and 0. The latter two results are disturbing.

TABLE 3
Computing the Sample Standard Deviation

Data Set	Certified Value	Excel	Minitab
Michelso:	0.0790105478190518	0.0790105482336454	0.0790105478190506
Pi Digits	2.86733906028871	2.86733906028871	2.86733906028871
Lew	277.332168044316	277.332168044316	277.332168044316
Lottery	291.699727470969	291.699727470969	291.699727470969
Mavro	0.000429123454003053	0.000429123453846293	0.000429123454003085
NumAcc1	1	1	1
NumAcc2	0.1	0.100000000000271	0.100000000000000
NumAcc3	0.1	0.107238052947636	0.100000000034925
NumAcc4	0.1	0.000000000000000	0.100000000558841

TABLE 4
LREs for Computing the Sample Standard Deviation

Data Set	Excel	Minitab	SAS	SPSS
Michelso:	8.3	13.8	13.8	12.4
Pi Digits	15	15	15	15
Lew	15	15	15	13.2
Lottery	15	15	15	15
Mavro	9.4	13.1	13.1	12.1
NumAcc1	15	15	15	15
NumAcc2	11.6	15	14.2	15
NumAcc3	1.1	9.5	9.5	9.5
NumAcc4	0	8.3	8.3	8.3

The performance of Minitab is reasonable for computing both the mean and standard deviation. As for Excel, the results suggest the need for incorporating a new algorithm to compute the standard deviation. Minitab performs well relative to SAS and SPSS in all cases.

As noted, these results mirror those of McCullough and Wilson (1999), suggesting that there have been no attempts at improving the algorithms of Excel. The deficiency in Excel's algorithm for the standard deviation leads MuCullough and Wilson to recommend against the use of Excel for data analysis. As an alternative in a teaching situation, this would provide an opportunity to have students program their own formula for standard deviation. It appears that Excel uses the formula:

$$\sqrt{\frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1}} \tag{2}$$

If the formula

$$\sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} \tag{3}$$

is programmed in Excel, the resulting LREs will be the same as those of Minitab.

Regression

The results for the coefficients in linear regression are shown in Table 5. Only the smallest LREs for the coefficients are shown in this table. This is consistent with McCullough's "a weakest link in the chain" principle: Use the minimum of the LREs for the statistic of interest. Excel does a reasonably good job of estimating the regression coefficients in all but one case. The Filip data set requires estimation of a 10^{th} degree polynomial regression, which results in a data set with extreme multicollinearity. Excel tries to estimate this equation but does not do a good job. Minitab prints out a warning about the high multicollinearity but goes ahead and computes the estimates and does a reasonably good job. SAS and SPSS do not even try to estimate the equation, pointing out the high multicollinearity. McCullough (1999) did note that the ORTHOREG command could be used in SAS to obtain a relatively accurate solution.

Excel has problems processing data sets with high multicollinearity. It would be better for Excel to compute some measure of multicollinearity and warn the user that estimates might be inaccurate or refuse to compute estimates if multicollinearity was extremely high rather than print out possibly inaccurate values. Given this caveat, Excel did reasonably well in all but the one case.

Minitab performs very well compared to SAS and SPSS in all cases.

TABLE 5
Minimum LREs for Estimates of Regression Coefficients

Data Set	Excel	Minitab	SAS	SPSS
Norris	12.1	12.2	12.3	12.3
Pontius	11.2	12.7	11.4	12.5
NoInt1	14.7	15	14.7	14.7
NoInt2	15	15	15	15
Filip	0	6.9*	ns	ns
Longley	7.4	12.7	8.6	12.1
Wampler1	6.6	9.6	8.3	6.6
Wampler2	9.6	12.6	10.0	9.7
Wampler3	6.6	9.3	7.0	7.4
Wampler4	6.6	8.7	7.0	7.4
Wampler5	6.6	6.8	7.0	5.8

* Prints out warning that predictors are highly correlated, but computes estimates.
ns Does not attempt to compute a solution.

Table 6 presents the LREs for the standard errors of the coefficients. The story here is essentially the same as for the coefficient estimates. Excel needs to check for extreme multicollinearity rather than trying to compute estimates. Otherwise it does reasonably well. Minitab does well overall.

To examine the deterioration of the estimates as the degree of multicollinearity increases, Table 7 shows the largest variance inflation factor for four of the data sets and also the LREs of the regression coefficient estimates for Minitab and Excel. Figure 1 shows the graph of LRE versus log (VIF). The LREs deteriorate much more rapidly for Excel than for Minitab. This relationship appears linear so a regression was fit using LRE as the dependent variable and log (VIF) as the independent variable. For Excel, the resulting equation is LRE = 10.5 - 0.264 log (VIF) while for Minitab it is LRE = 12.7 - 0.146 log (VIF). These equations can be used to predict the level where the two routines might be expected to fail. For example, with Excel, if an LRE of at least 5 is desired, log (VIF) should be no larger than 20.8. This would be a data set with an extremely high level of multicollinearity. However, since Excel does not include any multicollinearity diagnostics, it is impossible for the user to determine if a particular data set is a cause for concern.

ANOVA

Anova results are summarized using the F statistic value, since many intermediate values are used in computing this statistic. The LREs are summarized in Table 8. Excel handles the lower level difficulty data sets reasonably well, but performance on average and higher level

TABLE 6
Minimum LREs for Standard Errors of Regression Coefficients

Data Set	Excel	Minitab	SAS	SPSS
Norris	13.8	12.8	11.9	10.2
Pontius	14.3	11.8	9.2	8.9.
NoInt1	14.7	15	15	12.5
NoInt2	15	15	14.9	14.3
Filip	0	7.8*	ns	ns
Longley	8.5	14.3	10.3	13.3
Wampler1	7.1	15	15	6.6
Wampler2	11.8	15	15	9.7
Wampler3	11.2	13.5	10.9	10.6
Wampler4	11.2	13.7	11.5	10.8
Wampler5	11.2	13.7	11.5	10.8

* Prints out warning that predictors are highly correlated, but computes estimates.
ns Does not attempt to compute a solution

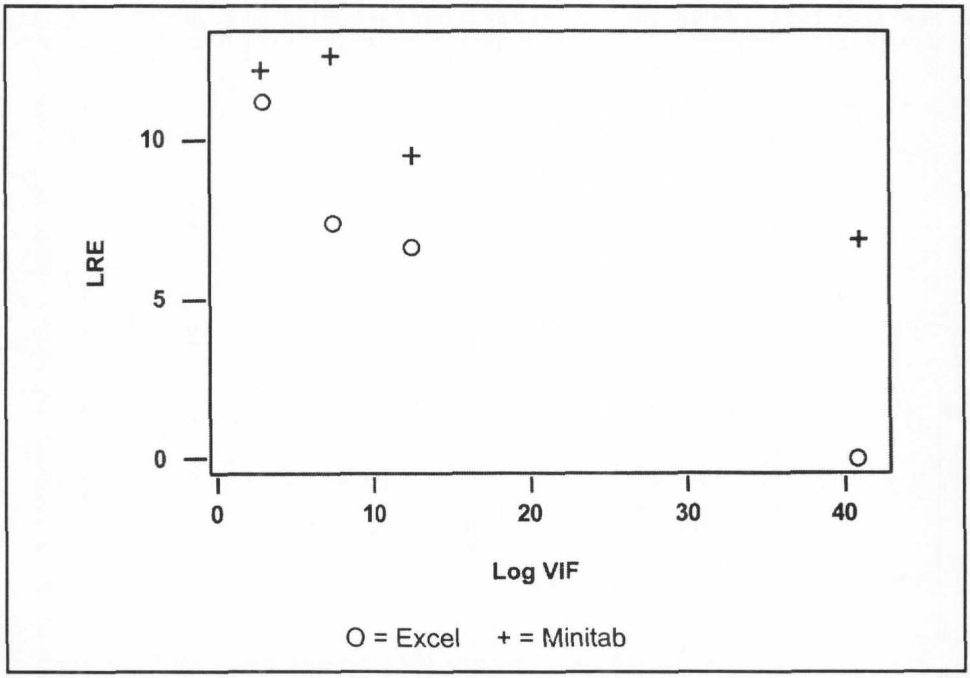

Figure 1. Graph of Log VIFs versus LREs for Minitab and Excel

TABLE 7
Minimum LREs and Maximum Variance Inflation Factors (VIFs)

Data Set	Maximum VIF	Excel	Minitab
Filip	5.5083E+17	0.0	6.9
Longley	1788.5	7.4	12.7
Norris	17.7	12.1	12.2
Wampler 1	261337.4	6.6	9.6

difficulty data sets is poor. Note that this is also true with SPSS and SAS. Revised algorithms for all three of these procedures should be implemented. It was more difficult to determine the LREs for Minitab since I could obtain the F statistic only to a limited number of significant digits. As a result, these values must be considered approximate. Given that, Minitab seems to do a good job with most of the data sets. The low value for SiRstv (3.4) is likely understated due to rounding. Three of the higher values (the 15s on SmLs04, SmLs05 and SmLs06) might be overstated due to rounding. Accuracy on the last three data sets drops for Minitab as well as for the other three packages although Minitab does better on SmLs07 and SmLs08.

TABLE 8
LREs for Anova F statistic

Data Set	Excel	Minitab	SAS	SPSS
SiRstv	7.4	3.4	8.3	9.6
SmLs01	14.3	15	13.3	15
SmLs02	12.5	15	11.4	15
SmLs03	12.6	15	11.8	12.7
AtmWtAg	1.8	3.7	0	miss
SmLs04	1.7	15	0	0
SmLs05	1.1	15	0	0
SmLs06	0*	15	0	0
SmLs07	0*	2.7	0	0
SmLs08	0	2.2	0	0
SmLs09	0*	0	0	0

* Computed negative F statistic value
miss Reported system missing-value

CONCLUSION

The accuracy of various statistical procedures in Excel 2000 and Minitab Version 13 was investigated using the StRD data sets. These results were compared to McCullough's (1999) results for SPSS and SAS. Results are as follows:

All packages compute the mean with acceptable accuracy. Excel has an inefficient algorithm to compute the standard deviation and will fail to compute an accurate value in some cases. It would be advisable to implement a different algorithm in Excel.

All four packages do a reasonable job of computing accurate values for regression coefficients and standard errors. Excel does fail when multicollinearity is extreme. SAS and SPSS may choose not to compute estimates in such extreme cases. Minitab produces a warning, but goes ahead and computes reasonably accurate estimates. The accuracy of Excel on less extreme data sets is not as high as that of the other three packages, but is reasonable.

All packages do reasonably well on the lower level difficulty data sets in the ANOVA category, but all have trouble on certain of the data sets in the average and higher level groups. Minitab seems to do the best job of all four packages in this category, although this result must be tempered somewhat, since enough digits of accuracy to make thorough comparisons for the F statistic were not available.

REFERENCES

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

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