Analyse-it

Evaluating the Numerical Accuracy of

Analyse-it for Microsoft Excel

This document describes the performance of Analyse-it for Microsoft Excel version 4.00 against the NIST StRD.

Analyse-it for Microsoft Excel is an add-in to Microsoft Excel that provides statistical analysis within Excel. It is designed to run on Microsoft Excel 2007, 2010, and 2013 on Microsoft Windows operating systems from Windows XP, Window Vista, Windows 7, 8, & 10 and Windows Server 2003, 2008, & 2012.

For more information see:

http://analyse-it.com/

7th August 2015

NIST StRD (Statistical Reference Datasets)

In response to industry concerns about the numerical accuracy of statistical software, the Statistical Engineering and Mathematical and Computational Sciences Divisions of NIST's Information Technology Laboratory developed datasets with certified values for a variety of statistical methods.

For more information about the datasets see:

http://www.itl.nist.gov/div898/strd/

The results obtained from statistical software packages can be compared against the certified values. The certified values are accurate to 15 significant digits and computed using ultra-high precision floating point arithmetic.

Most statistical packages use IEEE754 double precision (64bit) floating point arithmetic and due to finite precision, round-off and truncation errors involved in numerical operations, will be unable to obtain the exact certified value. Therefore, a good measure of the accuracy of a result *x* against the certified value *c*, can be calculated as the base-10 logarithm of the absolute value of the relative error:

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

if $c \neq 0$, *otherwise,*

$$\textbf{LRE} = -\log_{10}|x|$$

LRE is the number of significant digits in common with the certified value. Higher LRE values are better, and the maximum LRE obtainable is 15.

Performance benchmarks for the NIST StRD

We tested version 4.00 of Analyse-it using the NIST StRD on an Intel Xeon dual processor PC.

No statistical package achieves perfect accuracy for all the tests and no one package performs best for every test.

In the tests:

- **Analyse-it performed consistently and amongst the best on all tests.**
- **Analyse-it performed better than some of the more popular well-known statistical packages**.

Some developers of popular statistical software packages have published their own benchmarks against the NIST StRD, and some independent authors have also published benchmarks, see:

A comparative study of the reliability of nine statistical software packages,
Computational Statistics & Data Analysis, 51(8), 3811-3831,
Keeling, K. B., & Pavur, R. J.

On the Accuracy of Statistical Procedures in Microsoft Excel 97,
Computational Statistics and Data Analysis, July 1999, Volume 31, Number 1, pp 27-37,
McCullough, B.D. and Wilson, B.

Assessing the Reliability of Statistical Software: Part I,
The American Statistician, Volume 52, Number 4, pp 358-366,
McCullough, B.D.

Assessing the Reliability of Statistical Software: Part II,
The American Statistician, May 1999, Volume 53, Number 2, pp 149-159,
McCullough, B.D.

To download the Excel worksheets containing the Analyse-it analyses to perform the NIST StRD, and see comparisons against the published results for other packages, see:

http://analyse-it.com/support/Analyse-it-NIST-StRD-Validation.zip

Summary of results

The LRE obtained testing Analyse-it against the NIST StRD are summarized below.

Univariate summary statistics

The univariate tests consist of nine datasets classified by difficulty.

The mean and standard deviation were computed using the Distribution analysis and compared to the certified values.

The lag-1 autocorrelation is not computed by Analyse-it.

		LRE	
Test	**Difficulty**	**Mean**	**SD**
PiDigits	Lower	15.0	15.0
Lottery	Lower	15.0	15.0
Lew	Lower	15.0	15.0
Marvo	Lower	15.0	13.1
Michelson	Lower	15.0	13.8
NumAcc-1	Lower	15.0	15.0
NumAcc-2	Average	15.0	15.0
NumAcc-3	Average	15.0	9.5
NumAcc-4	Higher	15.0	8.3
Average		15.0	13.3
Minimum		15.0	8.3
Maximum		15.0	15.0

Analysis of variance

The analysis of variance tests consist of eleven datasets classified by difficulty.

The F statistic was computed using the Compare Groups – ANOVA analysis and compared to the certified value.

		LRE
Test	**Difficulty**	**F**
SiRstv	Lower	13.1
SmLs01	Lower	15.0
SmLs02	Lower	15.0
SmLs03	Lower	15.0
AtmWtAg	Average	10.2
SmLs04	Average	10.4
SmLs05	Average	10.2
SmLs06	Average	10.2
SmLs07*	Higher	4.4
SmLs08*	Higher	4.2
SmLs09*	Higher	4.2
Average		10.2 (12.4*)
Minimum		4.2 (10.2*)
Maximum		15.0

*Average/Minimum calculated excluding the marked tests.

NOTE: No statistical package has performed the Simon-Lesage tests 7, 8, and 9 (marked *) with more than 4.6 digits of accuracy, and although these programs reported 4.6 digits of accuracy for test 7 they performed markedly worse for tests 8 and 9. This is due to a flaw in the tests themselves rather than in the software packages, as the number 1,000,000,000,000.4 cannot be represented precisely using binary IEEE754 64bit double floating point representation. Instead it is represented as 1,000,000,000,000.4000244140625. Even simple summation of such a series of numbers leads to inaccuracy.

Linear regression

The linear regression tests consist of eleven datasets classified by difficulty.

The beta coefficients and standard error of the coefficients were computed using the Fit Model analysis and compared to the certified values. The minimum LRE value is reported for each analysis.

The R^2 statistic was computed and compared to the certified value. No R^2 value is computed for the no intercept models due to issues interpreting such a statistic.

The residual sum of square was computed and compared to the certified value.

		LRE			
Test	**Difficulty**	**Beta**	**Beta SE**	**Residual SS**	**R^2**
Norris	Lower	12.6	13.5	13.3	15.0
Pontius	Lower	12.4	12.7	12.4	15.0
NoInt1	Average	14.7	14.4	14.1	-
NoInt2	Average	15.0	15.0	14.6	-
Fillip	Higher	7.7	7.8	9.1	11.6
Longley	Higher	10.9	11.9	11.7	14.0
Wampler1	Higher	9.9	10.0	15.0	15.0
Wampler2	Higher	13.4	14.7	15.0	15.0
Wampler3	Higher	9.1	13.7	14.1	15.0
Wampler4	Higher	9.9	13.6	15.0	15.0
Wampler5	Higher	5.7	13.6	15.0	13.7
Average		11.0	12.8	13.6	14.4
Minimum		5.7	7.8	9.1	11.6
Maximum		15.0	15.0	15.0	15.0

NOTE: There may be slight variation in the LRE on repeated runs due to the use of available multiple processor cores when computing the QR decomposition of the matrix.