U. S. Department of the Interior
U. S. Geological Survey

Notes on Numerical Reliability of Several Statistical Analysis Programs

Open-File Report 99-95

Notes on Numerical Reliability of Several Statistical Analysis Programs

By J. M. Landwehr and G. D. Tasker

U.S. Geological Survey

Open-File Report 99-95

Reston, Virginia

1999

U. S. DEPARTMENT OF THE INTERIOR
Bruce Babbitt, *Secretary*

U. S. GEOLOGICAL SURVEY
Charles G. Groat, *Director*

For additional information write to:

Chief, Regional Hydrologic
Processes Project
U.S. Geological Survey
MS 431 - National Center
Reston, Virginia 20192 USA

Copies of this report can be purchased from:

U.S. Geological Survey
Branch of Information Services
Box 25286, Federal Center
Denver, Colorado 80225-0286 USA

CONTENTS

NOTES ON NUMERICAL RELIABILITY OF SEVERAL STATISTICAL ANALYSIS PROGRAMS

by J. M. Landwehr[1] and G. D. Tasker[2]

1. U. S. Geological Survey, 431 National Center, Reston, VA 20192
2. U. S. Geological Survey, 430 National Center, Reston, VA 20192

ABSTRACT

This report presents a benchmark analysis of several statistical analysis programs currently in use in the USGS. The benchmark consists of a comparison between the values provided by a statistical analysis program for variables in the reference data set ANASTY and their known or calculated theoretical values. The ANASTY data set is an amendment of the Wilkinson NASTY data set that has been used in the statistical literature to assess the reliability (computational correctness) of calculated analytical results.

INTRODUCTION

A universal expectation of any analyst using a statistical software program is that the package provides reliable (computationally correct) results. Conversely, the most dangerous situation occurs if a package provides statistics that look "reasonable" but are, in fact, grossly incorrect. Benchmarks are used to assess the reliability of statistical software. A benchmark consists of applying a suite of statistical analyses to various standard data sets for which the values of the statistics are known with great precision and assessing if the resulting values are in conformance. A discussion and review of how to perform such assessments can be found in Sawizki (1994a) and McCullough (1998).

McCullough made use of the Statistical Reference Datasets (StRDs) recently published by the National Institute of Standards and Technology (1998). To benchmark a software package by a large suite of analyses for all of the data sets in the StRDs is time consuming. McCullough (1998, p.358) pointed out, however, that "A well-known collection of simple tests is Wilkinson's (1985) *Statistics Quiz*, which presents a set of problems designed to uncover common flaws in statistical programs." In particular, Wilkinson proposed a data set called NASTY.DAT to test the limits of any analysis package, in which the variables are collinear and the difference in magnitude between variables was extreme but the magnitude of the observations for each variable was not unlike that found among common statistical problems. Sawitzki (1994b) reported on a joint effort by members of two working groups ("Computational Statistics" of the International Biometrical Society" and "Statistical Analysis Systems" of the GMDS) to apply the Wilkinson tests to nine data-analysis systems, including some running on multiple platforms. The group demonstrated performance difficulties with each system, even between platform implementations of the same package, but Sawitzki stressed (1994b, p.300) that "The presence of errors does not mean unusability for all purposes; the absence of error reports does not imply a recommendation." Rather, the work was a warning to researchers to

be aware of the potential for errors in their results and to raise the need for more quality control in commercial products.

Of the systems Sawitzki examined, two (SAS, Excel) currently are used widely within the USGS and one (S-PLUS) is expected to be used more frequently in the future This work examines the current versions of several programs used within the USGS with respect to an amended NASTY data set and a subset of the analyses proposed in the Wilkinson *Statistics Quiz* in order to examine the performance by and among these systems.

THE EXPERIMENT

A broad range of statistical programs is used within the USGS to address its diverse research and administrative objectives. The six software packages examined in this report include SAS v6.0701 and STATIT, running under UNIX on a Data General workstation; SAS v6.12, running under Windows NT4.0; Excel v.5.0a, MINITAB v10.51x, and KaleidaGraph v3.08d, running under Mac OS8.1; and S-PLUS v. 4.5, running under Windows NT4.[1]. Note that SAS, STATIT, MINITAB and S-PLUS are statistical analysis packages, *per se*, while Excel and KaleidaGraph have other primary functions, namely spreadsheet and graphing, respectively, which the statistical routines support.

Data analysis was performed using an amended NASTY data set. The Wilkinson data set NASTY contained the variables X, BIG, LITTLE, HUGE, TINY, ROUND, ZERO, and MISS. The variables BIG, LITTLE, HUGE, TINY, and ROUND are linear transforms of the variable X, but displaying the named characteristics, for the purpose of testing the calculating precision of a statistical routine. The variable ZERO is singular: all values are 0. The variable MISS has all missing values. We have extended NASTY by adding two other variables to the data set, namely XTRA and MIX. These two variables were added to examine the performance of each package in the case of extreme variation among the entries for a single variable (not unlike what is observed in the field for water-resource quantity and quality) and to see how a single missing entry for a variable might affect the reported results. We refer to this full set of 10 variables as ANASTY, as shown in Table 1.

The statistics computed for the variables in ANASTY include descriptive univariate statistics (mean, standard deviation, minimum, maximum, and median) shown in Table 2, as well as Pearson correlations among pairs of variables, shown in Table 3. Results are shown in the tables exactly as output, with no adjustment to display a comparable number of significant digits for each package. The values of various statistics for this data set are known theoretically or can be derived easily, so that the output can be checked to see how far the software results are from “truth”. Note that correct calculation with respect to this data set does not guarantee that the routines will always perform well, but incorrect results provide a warning about the limitations of the calculating routines.

Finally, we make the following disclaimer. All of these programs are quite extensive, with many control features: it is possible that some combination of features of which we are unaware might produce results other than what we present here. In the course of this work, we had no contact with nor did we receive advice from the technical or sales representatives from any of the companies that create or distribute these programs .

[1] We thank USGS colleague J.R.Slack for obtaining the S-PLUS results for us on his workstation.

OBSERVATIONS

Several surprising issues surfaced in the course of the study. These matters were not a concern *a priori*, but do bear on the conclusions drawn from the results.

First, the internal precision with which the computations are done is critical to getting a "correct" value, but it is also necessary to be sensitve to the format by which the results are reported. Three of the packages do not allow format control over the displayed results (SAS, STATIT, and MINITAB) whereas three do allow great latitude in display (Excel, KaleidaGraph, and S-PLUS). When formats are prespecified, it is not always clear from the number of significant digits of the printed outputs if the calculations were incorrect or if just the results were badly rounded to a limiting degree of imprecision. For example, MINITAB provided a median value of 9, rather than 8.75, for the variable MIX, which may reflect only rounding imprecision, but might be a calculation error. But when control over the format of displayed values is allowed, the user must make a careful decision about how many significant digits should be seen in the case of each statistic balanced against what is computationally possible by the program, which may not be known.

Second, with respect to computing the descriptive univariate statistics, the routine PROC_CORR from SAS (on both the UNIX and Windows NT platforms) and S-PLUS provided values that consistently reproduced the known values, but S-PLUS does require some format display decisions. The PROC_UNIVARIATE routine of SAS (on both platforms), STATIT, Excel, MINITAB and KaleidaGraph all had some difficulty in producing correct results for some variables, particularly for the variables BIG and LITTLE, possibly reflecting imprecision in the internal representation and calculation of these numbers. However, the SAS call to PROC_UNIVARIATE running under WindowsNT did provide a "WARNING" for the variables BIG and LITTLE on the log sheet (not with the other output) that the range was too small relative to the mean of the series, and that numerical inaccuracies may occur so that the user should recode the data if possible, although no such warning was provided by SAS running under the Data General UNIX.

Third, a statistical package may not be consistent in its output! We observed four instances of this. First, the PROC_UNIVARIATE routine of SAS provided different results than PROC_CORR. For example, for the variable BIG, PROC_UNIVARIATE gives a value of "0" for the standard deviation, whereas PROC_CORR gave the correct value of 2.738613. Second, as noted above, the SAS package running under Windows NT provided a warning which was not produced by SAS running on the Data General UNIX, although the outputs given by both routines are otherwise identical. Third, KaleidaGraph provides the Pearson correlation coefficient when fitting a line to an x-y plot, but the estimated value of the correlation differed depending on the plotting order of the variables; that is, (x,y) and (y,x) analyses did not always yield the same R value. Fourth, in the case of missing observations, S-PLUS provided one set of estimates for correlations with the "omit" option and another with the "available" option. To be sure, one can call up an explanation that indicates each option is using a different algorithm and different number of observations to compute the reported statistic, but in many cases, the analyst has no standard by which to decide which option to choose.

The fourth observation we made was with respect to the calculations of Pearson correlations: it is necessary to know how missing values for the variables in a data set are treated in order to even know if the results are appropriate to answer the question the analyst wishes to ask, much less to assess the computational correctness of the results. Most frequently, one wishes to estimate the correlation between each pair of variables in a data set, making use of all observations common to each pair. Results with this objective are shown in Table 3. The correlation matrix that is so defined, however, may not itself be statistically well-behaved, that is, it may not be positive definite, which may or may not be important depending on the objective of the analysis. In contrast, in some situations the analyst may want an estimate of the cross-correlations among several variables in a data set but only for those cases which have observations for all of the variables of interest. For example, within a region, station records for different stations may have missing data for different periods due to some operational difficulty with the gage at each particular station, but if a peak or peaks of record are missing from some gages and not from others, that is, if the representation of the occurrence of extremes was inconsistent among the regional station records, it could be misleading to compare by-variable correlations between station pairs within the region.

The default in the PROC_CORR routine of SAS is to compute correlations by variable pair for all variables in the specified data set. That is, SAS computes a correlation coefficient based on the mutual observations between each pair of variables and then provides the estimated statistic for each pair in the output matrix, as well as the number of observations used in its computation and its respective significance level. Should the analyst want the correlations computed on just the set of observations common to all variables, however, then the "NOMISS" option can be chosen.

MINITAB also provides a by-variable pair calculation in response to the CORR command issued for a set of variables, but it does not display the number of observations used for each statistic. Should the analyst want the correlations on just the common set of observations, then the data set must be amended manually before submitting it to CORR.

Excel and KaleidaGraph provide an estimate of the Pearson correlation but the command is issued only for each pair, and not for an entire set of variables. Consequently. the statistic is computed only in a by-pair mode, although the data set can be amended manually to reflect any particular subset pairing.

STATIT provides correlations only for the set of observations common to all variables in a data set. In order to obtain the by-pair results for the ANASTY data set, as shown in Table 3, several subsets of variables had to be submitted separately to STATIT.

S-PLUS provides several options for obtaining the correlations of variables in a data set. The "fail" option will not compute a correlation matrix for a data set if any data is missing. The "omit" option is equivalent to the STATIT function in that it omits all observations (rows) for which there is a missing value so that the correlations are estimated only for the common set of observations. The "available" option uses all values for all variables, but with a special treatment of the covariance matrix to account for missing values, so that the estimated correlation statistic for pairs in which at least one of the variables has a missing observation is different than the statistic that is computed using just the data common to the variable pair. As in the case of STATIT, to use S-PLUS to obtain the by-variable pair correlations for all of the

ANASTY variables as shown in Table 3, we needed to submit multiple data subsets for analysis. This can be cumbersome with a large data set.

Table 3 shows that the PROC_CORR routine of SAS, S-PLUS and Excel reproduced the theoretical values, KaleidaGraph had some computational difficulties with the variable LITTLE, whereas STATIT and MINITAB had difficulty with both LITTLE and BIG.

CONCLUSIONS

We stress that this particular benchmark study only illustrates the performance of the statistical software examined, repeating the comment by Sawitzki (1994b), that "the presence of errors does not imply unusability for all purposes; the absence of error reports does not imply a recommendation". We did observe that several analysis packages did have more difficulty in providing computationally precise and/or correct values than did others, and some were more cumbersome to use in obtaining specific statistics. Also, the users must be "educated consumers" when using statistical software; that is they must be aware of what question they specifically want answered by the analysis ,as well as how a specific package is computing the answer it provides. Finally, we suggest that the ANASTY data set be used to run analyses to acquaint a researcher with characteristic behavior of any statistical analysis package before they use that software to analyze their specific study data sets.

REFERENCES CITED

McCullough, B., 1998: "Assessing the Reliability of Statistical Software: Part I", THE AMERICAN STATISTICIAN, v.52, no. 4, p.358-366.

National Institute of Standards and Technology, 1998, "Statistical Reference Datasets", http://www.nist.gov/itl/div898/strd.

Sawitzki, G., 1994a, "Testing numerical reliability of data analysis systems," Computational Statistics and Data Analysis, v. 18, p.269-286.

Sawitzki, G., 1994b, "Report on the reliability of data analysis systems," Computational Statistics and Data Analysis, v. 18, p.289-301.

Wilkinson, L., 1985, Statistics Quiz, Systat Inc., Evanston, IL. (Also, available via internet at http://www.tspintl.com/products/tsp/benchmarks/wilk.txt)

Table 1. ANASTY, or the amended NASTY dataset with variables XTRA and MIX added.

VARIABLE NAME	X	BIG	LITTLE	HUGE	TINY	ROUND	XTRA	MIX	ZERO	MISS
ONE	1	99999991	0.99999991	1E+12	1E-12	0.5	1	9	0	*
TWO	2	99999992	0.99999992	2E+12	2E-12	1.5	2	0	0	*
THREE	3	99999993	0.99999993	3E+12	3E-12	2.5	3000	*	0	*
FOUR	4	99999994	0.99999994	4E+12	4E-12	3.5	4	99999999	0	*
FIVE	5	99999995	0.99999995	5E+12	5E-12	4.5	5	0.99999999	0	*
SIX	6	99999996	0.99999996	6E+12	6E-12	5.5	6	9E+12	0	*
SEVEN	7	99999997	0.99999997	7E+12	7E-12	6.5	7	9E-12	0	*
EIGHT	8	99999998	0.99999998	8E+12	8E-12	7.5	8	8.5	0	*
NINE	9	99999999	0.99999999	9E+12	9E-12	8.5	9	9.4	0	*

* indicates missing value.

Table 2. Theoretical values for descriptive univariate statistics for ANASTY and the computed results from several programs.

Statistic for Variable:	Mean	Standard Deviation	Minimum	Maximum	Median
Theoretical					
X	5	2.738612788...	1	9	5
BIG	99999995	2.738612788...	99999991	99999999	99999995
LITTLE	0.99999995	2.738612788E-8	0.99999991	0.99999999	0.99999995
HUGE	5E12	2.738612788E12	1E12	9E12	5E12
TINY	5E-12	2.738612788E-12	1E-12	9E-12	5E-12
ROUND	4.500000	2.738612788...	0.500000	8.500000	4.5
XTRA	338	998.25347...	1	3000	6
MIX	1.1250125E12	3.18197E12	0	9E12	8.75
ZERO	0	0	0	0	0
MISSING	na	na	na	na	na
SAS PROC_UNIVARIATE for DG v6.0701 and for WindowsNT v6.12					
X	5	2.738613	1	9	5
BIG	99999995	0	99999991	99999999	99999995
LITTLE	1	0	1	1	1
HUGE	5E12	2.739E12	1E12	9E12	5E12
TINY	5E-12	2.74E-12	1E-12	9E-12	5E-12
ROUND	4.5	2.738613	0.5	8.5	4.5
XTRA	338	998.2535	1	3000	6
MIX	1.125E12	3.182E12	0	9E12	8.75
ZERO	0	0	0	0	0
MISSING	.	.	.	.	.
SAS PROC_CORR for DG v6.0701 and for WindowsNT v6.12					
X	5.000000	2.738613	1.000000	9.000000	*
BIG	99999995	2.738613	99999991	99999999	*
LITTLE	1.000000	2.7386128E-8	1.000000	1.000000	*
HUGE	5E12	2.7386128E12	1E12	9E12	*
TINY	5E-12	2.738613E-12	1E-12	9E-12	*
ROUND	4.500000	2.738613	0.500000	8.500000	*
XTRA	338.000000	998.253475	1	3000	*
MIX	1.1250125E12	3.1819755E12	0	9E12	*
ZERO	0	0	0	0	*
MISSING	*	*	*	*	*

Table 2. Theoretical values for descriptive univariate statistics for ANASTY and the computed results from several programs.--*Continued*

Statistic for Variable:	Mean	Standard Deviation	Minimum	Maximum	Median
STATIT DG					
X	5	2.738613	1	9	5
BIG	1E+8	4.21637	1E+8	1E+8	1E+8
LITTLE	0.9999999	3.582E-8	0.9999999	1	0.9999999
HUGE	5E+12	2.739E+12	1E+12	9E+12	5E+12
TINY	5E-12	2.739E-12	1E-12	9E-12	5E-12
ROUND	4.5	2.738613	0.5	8.5	4.5
XTRA	338	998.2535	1	3000	6
MIX	1.125E+12	3.182E+12	0	9E+12	8.75
ZERO	0	0	0	0	0
MISSING	*	*	*	*	*
Excel PowerPC Mac v5.0a					
X	5	2.73861279	1	9	5
BIG	99999995	2.44948974	99999991	99999999	99999995
LITTLE	0.99999995	2.9802E-08	0.99999991	0.99999999	0.99999995
HUGE	5E+12	2.7386E+12	1E+12	9E+12	5E+12
TINY	5E-12	2.7386E-12	1E-12	9E-12	5E-12
ROUND	4.5	2.73861279	0.5	8.5	4.5
XTRA	338	998.253475	1	30000	6
MIX	1.125E+12	3.182E+12	0	9E+12	8.75
ZERO	0	0	0	0	0
MISSING	*	*	0	0	*
MINITAB PowerPC Mac v10.51x					
X	5.000	2.739	1.000	9.000	5.000
BIG	99999992	4	99999992	1.00E+08	99999992
LITTLE	1.0000	0.0000	1.0000	1.0000	1.0000
HUGE	5.00E+12	2.74E+12	1.00E+12	9.00E+12	5.00E+12
TINY	0.00000	0.00000	0.00000	0.00000	0.00000
ROUND	4.500	2.739	0.500	8.500	4.5
XTRA	338	998	1	3000	6
MIX	1.13E+12	3.18E+12	0	9.00E+12	9
ZERO	0.00000	0.00000	0.00000	0.00000	0.00000
MISSING	*	*	*	*	*

Table 2. Theoretical values for descriptive univariate statistics for ANASTY and the computed results from several programs.--*Continued*

Statistic for Variable:	Mean	Standard Deviation	Minimum	Maximum	Median
KaleidaGraph PowerPC Mac v3.08d					
X	5	2.73861289	1	9	5
BIG	99999991.99	2.35702252	99999991.99	100000000	99999991.99
LITTLE	0.999999940	0.0000000289626509	0.9999998807907	1	0.9999999403954
HUGE	4999999913984	2738612862976	999999995904	9000000159744	4999999913984
TINY	0.000000000005	0.0000000000027386	0.000000000001	0.000000000009	0.000000000005
ROUND	4.5	2.73861289	0.5	8.5	4.5
XTRA	338	998.253479...	1	3000	6
MIX	1125012471808	3181975437312	0	9000000159744	8.75
ZERO	0	0	0	0	0
MISSING	0	0	0	0	0
S-PLUS WindowsNT v4.5					
X	5.0000	2.73861	1.00000	9.00000	5.0000
BIG	99999995.00000	2.73861	99999991.00000	99999999.00000	99999995.00000
LITTLE	0.999999950000000	.0000000273861279	0.999999910000000	0.999999990000000	0.999999950000000
HUGE	5.00000e+12	2.73861e+12	1.00000e+12	9.00000e+12	5.00000e+12
TINY	5.00000e-12	2.73861e-12	1.00000e-12	9.00000e-12	5.00000e-12
ROUND	4.50000	2.73861	0.50000	8.50000	4.50000
XTRA	338	998	1	3000	6
MIX	1125012500003.36	3181975464767.79	0.00	900000000000.00	8.75
ZERO	0.00	0.00	0.00	0.00	0.00
MISSING	*	*	*	*	*

na = Not apply
*** = Not given**

Table 3. Pair-wise Pearson correlations for ANASTY, for theortetical values and the results given by statisical routines.
[Correlations use nine pairs of observations, except those with MIX use only eight observations. Note that SAS, and MINITAB produce pair-wise correlations with single command for the dataset, Excel and KaleidaGraph require multiple commands by variable-pair, and STATIT and S-PLUS require multiple commands by subsets of variables. The "available" option for S-PLUS produces pair-wise statistics with a single command for the dataset, but introduces a potentially unexpected "correction" in the case of a variable with a missing observation.]

Variable:	**X**	**BIG**	**LITTLE**	**HUGE**	**TINY**	**ROUND**	**XTRA**	**MIX**	**ZERO**	**MISS**
Theoretical										
X	1	1	1	1	1	1	-0.2713...	0.10762..	na	na
BIG	1	1	1	1	1	1	-0.2713...	0.10762..	na	na
LITTLE	1	1	1	1	1	1	-0.2713...	0.10762..	na	na
HUGE	1	1	1	1	1	1	-0.2713...	0.10762..	na	na
TINY	1	1	1	1	1	1	-0.2713...	0.10762..	na	na
ROUND	1	1	1	1	1	1	-0.2713...	0.10762..	na	na
XTRA	-0.2713...	-0.2713...	-0.2713...	-0.2713...	-0.2713...	-0.2713...	1	0.10762	na	na
MIX	0.10762..	0.10762..	0.10762..	0.10762..	0.10762..	0.10762..	0.10762	1	na	na
ZERO	na	na	na	na	na	na	na	na	1	na
MISS	na	na	na	na	na	na	na	na	na	na
SAS PROC_CORR for DG V6.0701 and for Window NT v6.12										
X	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	-0.27132	0.10762	.	*
BIG	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	-0.27132	0.10762	.	*
LITTLE	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	-0.27132	0.10762	.	*
HUGE	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	-0.27132	0.10762	.	*
TINY	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	-0.27132	0.10762	.	*
ROUND	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	-0.27132	0.10762	.	*
XTRA	-0.27132	-0.27132	-0.27132	-0.27132	-0.27132	-0.27132	1	0.10762	.	
MIX	0.10762	0.10762	0.10762	0.10762	0.10762	0.10762	0.10762	1.00000	.	*
ZERO	.	.	.	.	.	.	.	.	.	*
MISS	*	*	*	*	*	*		*	*	*

Table 3. Pair-wise Pearson correlations for ANASTY, for theortetical values and the results given by statisical routines.
[Correlations use nine pairs of observations, except those with MIX use only eight observations. Note that SAS, and MINITAB produce pair-wise correlations with single command for the dataset, Excel and KaleidaGraph require multiple commands by variable-pair, and STATIT and S-PLUS require multiple commands by subsets of variables. The "available" option for S-PLUS produces pair-wise statistics with a single command for the dataset, but introduces a potentially unexpected "correction" in the case of a variable with a missing observation.] (*Continued*)

Variable:	**X**	**BIG**	**LITTLE**	**HUGE**	**TINY**	**ROUND**	**XTRA**	**MIX**	**ZERO**	**MISS**
STATIT DG										
X	1.00000	0.86603	0.83551	1.00000	1.00000	1.00000	-0.27132	0.10762	99.00000	*
BIG	0.86603	1.00000	0.61394	0.86603	0.86603	0.86603	-0.27132	0.37796	99.00000	*
LITTLE	0.83551	0.61394	1.00000	0.83551	0.83551	0.83551	-0.27132	-0.07881	99.00000	*
HUGE	1.00000	0.86603	0.83551	1.00000	1.00000	1.00000	-0.27132	0.10762	99.00000	*
TINY	1.00000	0.86603	0.83551	1.00000	1.00000	1.00000	-0.27132	0.10762	99.00000	*
ROUND	1.00000	0.86603	0.83551	1.00000	1.00000	1.00000	-0.27132	0.10762	99.00000	*
XTRA	-0.27132	-0.27132	-0.27132	-0.27132	-0.27132	-0.27132	1.00000	0.10762	99.00000	
MIX	0.10762	0.37796	-0.07881	0.10762	0.10762	0.10762	0.10762	1.00000	99.00000	*
ZERO	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	99.00000	1.00000	*
MISS	*	*	*	*	*	*	*	*	*	*
Excel PowerPC Mac V5.0a										
X	1	1	1	1	1	1	-0.2713227	0.10762258	*	*
BIG	1	1	1	1	1	1	-0.2713227	0.10762258	*	*
LITTLE	1	1	1	1	1	1	-0.2713227	0.10762258	*	*
HUGE	1	1	1	1	1	1	-0.2713227	0.10762258	*	*
TINY	1	1	1	1	1	1	-0.2713227	0.10762258	*	*
ROUND	1	1	1	1	1	1	-0.2713227	0.10762258	*	*
XTRA	-0.27132 27	-0.27132 27	-0.27132 27	-0.27132 27	-0.27132 27	-0.27132 27	1	0.10762258	*	*
MIX	0.1076225 8	0.1076225 8	0.1076225 8	0.1076225 8	0.1076225 8	0.1076225 8	0.10762258	1	*	*
ZERO	*	*	*	*	*	*	*	*	*	*
MISS	*	*	*	*	*	*	*	*	*	*

Table 3. Pair-wise Pearson correlations for ANASTY, for theortetical values and the results given by statisical routines.
[Correlations use nine pairs of observations, except those with MIX use only eight observations. Note that SAS, and MINITAB produce pair-wise correlations with single command for the dataset, Excel and KaleidaGraph require multiple commands by variable-pair, and STATIT and S-PLUS require multiple commands by subsets of variables. The "available" option for S-PLUS produces pair-wise statistics with a single command for the dataset, but introduces a potentially unexpected "correction" in the case of a variable with a missing observation.] (*Continued*)

Variable:	**X**	**BIG**	**LITTLE**	**HUGE**	**TINY**	**ROUND**	**XTRA**	**MIX**	**ZERO**	**MISS**
MINITAB PowerPC Mac V10.51x										
X	*	*	*	*	*	*	*	*	*	*
BIG	0.866	*	*	*	*	*	*	*	*	*
LITTLE	0.836	0.614	*	*	*	*	*	*	*	*
HUGE	1.000	0.866	0.836	*	*	*	*	*	*	*
TINY	1.000	0.866	0.836	1	*	*	*	*	*	*
ROUND	1.000	0.866	0.836	1	1	*	*	*	*	*
XTRA	-0.271	-0.314	-0.067	-0.271	-0.271	-0.271	*	*	*	*
MIX	0.108	0.378	-0.079	0.108	0.108	0.108	0.108	*	*	*
ZERO	*	*	*	*	*	*	*	*	*	*
MISS	*	*	*	*	*	*	*	*	*	*
KaleidoGraph PowerPC Mac V3.08d										
X	*	1	1	1	1	1	0.27132	0.10762	0	*
BIG	1	*	1	1	1	1	0.27132	0.10762	0	*
LITTLE	1	0.87896	*	1	1	1	0.27132	0.10762	0	*
HUGE	1	1	0.7082	*	1	1	0.27132	0.10762	0	*
TINY	1	1	0.88921	1	*	1	0.27132	0.10762	0	*
ROUND	1	1	0.7082	1	1	*	0.27132	0.10762	0	*
XTRA	0.27132	1	-0.12964	0.27132	0.27132	0.27132	*	0.10762	0	*
MIX	0.10762	0	0.40825	0.10762	0.10762	0.10762	0.10762	*	0	*
ZERO	*	*	*	*	*	*	*	*	*	*
MISS	*	*	*	*	*	*		*	*	*

Table 3. Pair-wise Pearson correlations for ANASTY, for theortetical values and the results given by statisical routines.
[Correlations use nine pairs of observations, except those with MIX use only eight observations. Note that SAS, and MINITAB produce pair-wise correlations with single command for the dataset, Excel and KaleidaGraph require multiple commands by variable-pair, and STATIT and S-PLUS require multiple commands by subsets of variables. The "available" option for S-PLUS produces pair-wise statistics with a single command for the dataset, but introduces a potentially unexpected "correction" in the case of a variable with a missing observation.] (*Continued*)

Variable:	**X**	**BIG**	**LITTLE**	**HUGE**	**TINY**	**ROUND**	**XTRA**	**MIX**	**ZERO**	**MISS**
MINITAB PowerPC Mac V10.51x										
X	*	*	*	*	*	*	*	*	*	*
BIG	0.866	*	*	*	*	*	*	*	*	*
LITTLE	0.836	0.614	*	*	*	*	*	*	*	*
HUGE	1.000	0.866	0.836	*	*	*	*	*	*	*
TINY	1.000	0.866	0.836	1	*	*	*	*	*	*
ROUND	1.000	0.866	0.836	1	1	*	*	*	*	*
XTRA	-0.271	-0.314	-0.067	-0.271	-0.271	-0.271	*	*	*	*
MIX	0.108	0.378	-0.079	0.108	0.108	0.108	0.108	*	*	*
ZERO	*	*	*	*	*	*	*	*	*	*
MISS	*	*	*	*	*	*	*	*	*	*
KaleidoGraph PowerPC Mac V3.08d										
X	*	1	1	1	1	1	0.27132	0.10762	0	*
BIG	1	*	1	1	1	1	0.27132	0.10762	0	*
LITTLE	1	0.87896	*	1	1	1	0.27132	0.10762	0	*
HUGE	1	1	0.7082	*	1	1	0.27132	0.10762	0	*
TINY	1	1	0.88921	1	*	1	0.27132	0.10762	0	*
ROUND	1	1	0.7082	1	1	*	0.27132	0.10762	0	*
XTRA	0.27132	1	-0.12964	0.27132	0.27132	0.27132	*	0.10762	0	*
MIX	0.10762	0	0.40825	0.10762	0.10762	0.10762	0.10762	*	0	*
ZERO	*	*	*	*	*	*	*	*	*	*
MISS	*	*	*	*	*	*		*	*	*

Table 3. Pair-wise Pearson correlations for ANASTY, for theortetical values and the results given by statisical routines. [Correlations use nine pairs of observations, except those with MIX use only eight observations. Note that SAS, and MINITAB produce pair-wise correlations with single command for the dataset, Excel and KaleidaGraph require multiple commands by variable-pair, and STATIT and S-PLUS require multiple commands by subsets of variables. The "available" option for S-PLUS produces pair-wise statistics with a single command for the dataset, but introduces a potentially unexpected "correction" in the case of a variable with a missing observation.] (*Continued)*

Variable:	**X**	**BIG**	**LITTLE**	**HUGE**	**TINY**	**ROUND**	**XTRA**	**MIX**	**ZERO**	**MISS**
S-PLUS WindowsNT v4.5										
X	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	-0.2713227	0.1076226	*	*
BIG	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	-0.2713227	0.1076226	*	*
LITTLE	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	-0.2713227	0.1076226	*	*
HUGE	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	-0.2713227	0.1076226	*	*
TINY	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	-0.2713227	0.1076226	*	*
ROUND	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	1.0000000	-0.2713227	0.1076226	*	*
XTRA	-0.27132	-0.27132	-0.27132	-0.27132	-0.27132	-0.27132	1.0000000	0.1076226	*	*
	27	27	27	27	27	27				
MIX	0.1076226	0.1076226	0.1076226	0.1076226	0.1076226	0.1076226	0.1076226	1.000000	*	*
ZERO	*	*	*	*	*	*	*	*	*	*
MISS	*	*	*	*	*	*	*	*	*	*
S-PLUS WindowsNT v4.5 using the "available" option										
X	1	1	1	1	1	1	-0.2713227	0.1106548	*	*
BIG	1	1	1	1	1	1	-0.2713227	0.1106548	*	*
LITTLE	1	1	1	1	1	1	-0.2713227	0.1106548	*	*
HUGE	1	1	1	1	1	1	-0.2713227	0.1106548	*	*
TINY	1	1	1	1	1	1	-0.2713227	0.1106548	*	*
ROUND	1	1	1	1	1	1	-0.2713227	0.1106548	*	*
XTRA	-0.2713	-0.27132	-0.27132	-0.27132	-0.27132	-0.27132	1	0.003035708	*	*
	227	27	27	27	27	27				
MIX	0.1106548	0.1106548	0.1106548	0.1106548	0.1106548	0.1106548	0.003035708	1	*	*
ZERO	*	*	*	*	*	*	*	*	*	*
MISS	*	*	*	*	*	*	*	*	*	*