# Assessing the Numerical Accuracy of SAS Software

Statistical R&D Staff
SAS Institute Inc., Cary, NC 27513

## Introduction

Statisticians and data analysts have long been concerned with the numerical accuracy of statistical algorithms provided by commercial software. Statistical software companies perform extensive testing of their software to assure that the implemented algorithms compute accurate results.

The National Institute of Standards and Technology (NIST) has assembled a challenging array of tests for assessing the computational accuracy of statistical software.  Recently, reviewers of software numerical accuracy have taken advantage of these test cases in their work, for example, McCullough (1998,1999). This document provides an overview of the NIST tests, describes relevant SAS procedures for analyzing the NIST data sets, and summarizes the performance of the SAS System for these and other tests of numerical accuracy.

## NIST Statistical Reference Datasets

The NIST *Statistical Reference Datasets* (StRD) tests establish benchmarks in four general areas: univariate statistics, analysis of variance, linear regression, and nonlinear regression. The scenarios are designed to test the numerical accuracy of statistical algorithms and represent extreme cases of ill-conditioned data over a range of difficulties.

Sources of the StRD tests are published research, analyses by NIST consultants, and artificial data sets. Although these artificial data sets were created to imitate real-world situations and not just to break a specific algorithm, they are typically more pathological and ill-conditioned than the data most statisticians and analysts will encounter in practice.

For more information on the NIST StRD tests, downloadable data sets, and certified values that serve as "best-available" solutions, visit the NIST StRD web site at:

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

## Accuracy in SAS Procedures

This section discusses the basic statistical procedures in the SAS System for the areas covered by the StRD tests and describes their performance for the tests. These tests were performed using SAS Software, Release 6.12 on the Windows NT platform.  The tests were also executed using the Version 8.02 release.

Numerical accuracy is measured by the number of correct significant digits, computed as the base-10 logarithm of the relative error

$$LRE = -\log_{10}[|e\text{-}c|/|c|]$$

Where $e$ is the estimated value and $c$ is the correct value.  McCullough (1998) provides further details on measuring numerical accuracy.

### Univariate Statistics

The UNIVARIATE procedure is a primary tool for computing univariate summary statistics in the SAS System.  PROC UNIVARIATE produces very accurate estimates of the mean and standard deviation for the StRD data sets.

### ANOVA and Linear Regression Background

The standard SAS/STAT software regression procedures (REG, ANOVA, and GLM) are very accurate for most regression analyses. However, if you have very ill-conditioned data, these procedures can produce estimates that yield an error sum of squares very close to the minimum but still different from the exact least-squares estimates.  When this occurs, it is usually associated with very large standard errors for the estimates. See Thisted (1988) for further information regarding numerical computing issues for statistical computations.

Many procedures issue a note to the SAS log when they encounter data that are potentially numerically sensitive.  For example, the GLM procedure writes the following note when it determines that the data might be problematic:

> "WARNING: Due to the range of values for Response numerical inaccuracies may occur. Recode if possible."

In addition, the REG procedure also issues a warning to the SAS listing when it encounters singularities.

The ORTHOREG procedure in SAS/STAT software was designed to handle numerically sensitive data and is the recommended procedure for regression and analysis of variance when your data are ill-conditioned.  Instead of collecting crossproducts,

PROC ORTHOREG uses Gentleman-Givens transformations to compute the QR decomposition of the data matrix, with special care for scaling (Gentleman 1972, 1973). This method has the advantage over other orthogonalization methods of not requiring the data matrix to be stored in memory.

## ANOVA and Linear Regression Results

The REG procedure performed very well when exercised with the StRD data sets, except for the case of the FILIPPEL data. For these data, PROC REG incorrectly detects some singularity, sets some parameters estimates to zero, and prints a message to the listing effectively describing what has occurred.

In Release 6.12, the ANOVA procedure produces accurate estimates of the $F$ statistic for the less difficult ANOVA tests, but it does not perform as well for the difficult problems. The Version 8.02 Release of PROC ANOVA has been updated to handle better those data with small coefficients of variation. It computes the $F$-statistic with greater accuracy (10 digits for average problems and 4 digits for hard problems).

For these more difficult problems where the data are extremely ill-conditioned, the ORTHOREG procedure provides a suitable alternative. PROC ORTHOREG produces estimates of the $F$-statistic that are accurate to at least 10 digits on average problems and 4 digits for the three hard problems.

Table 1 displays the results for the StRD Analysis of Variance tests for the PROC ANOVA using both Release 6.12 and Version 8.02 and PROC ORTHOREG using Version 8.02. PROC ORTHOREG computes similar results in Release 6.12 using the Mean Square Error statistic.

| | Procedure and Version of SAS | | |
|---|---|---|---|
| | **ANOVA** | **ORTHOREG** | **ANOVA** |
| Test | 6.12 | 8.2 | 8.2 |
| SiRstv | 8.3 | 13.1 | 12.7 |
| SmnLsg01 | 13.3 | 14.8 | 15 |
| SmnLsg02 | 11.4 | 14.5 | 13.9 |
| SmnLsg03 | 11.8 | 14.4 | 12.7 |
| AtmWtAg | 0.9 | 10.2 | 8.8 |
| SmnLsg04 | 0.8 | 10.4 | 10.4 |
| SmnLsg05 | 0 | 10.2 | 10.2 |
| SmnLsg06 | 0 | 10.2 | 10.2 |
| SmnLsg07 | 0 | 4.4 | 4.4 |
| SmnLsg08.0 | 0 | 4.2 | 4.2 |
| SmnLsg09 | 0 | 4.2 | 4.2 |

**Table 1:** Significant Digits of Accuracy for the StRD Analysis of Variance Tests

## Nonlinear Regression Background

The NLIN procedure is the primary procedure for fitting nonlinear regression models in SAS/STAT software. By default, the NLIN procedure computes analytic derivatives and uses the Gauss-Newton method with a convergence criterion of $10^{-5}$.

Although PROC NLIN is typically used for fitting nonlinear regression models, you can also fit nonlinear models using the MODEL procedure in SAS/ETS software and the NLP procedure in SAS/OR software. PROC MODEL analyzes models in which the relationships among the variables comprise a system of one or more nonlinear equations. Primary uses of the MODEL procedure are estimation, simulation, and forecasting of nonlinear simultaneous equation models. PROC NLP offers a set of optimization techniques for minimizing or maximizing continuous nonlinear functions with lower and upper bounds, linear and nonlinear, and equality and inequality constraints.

The nonlinear regression problems require a little more effort when you want to exercise them with the StRD data sets. For each nonlinear model to be analyzed, you must specify the model and the names and starting values of the parameters to be estimated. Moreover, nonlinear estimation requires an iterative process to converge to an accurate solution. Results of nonlinear estimation can therefore vary widely depending on convergence tolerance, type of solution, convergence criteria, and the choice of numeric or analytic derivatives.

## Nonlinear Regression Results

For the StRD data sets, the NLIN procedure in Release 6.12 converges to a solution from a starting point far from the solution in 20 of the 27 nonlinear regression StRD problems. However, PROC NLIN converges to a solution for 26 of the 27 nonlinear regression problems using a starting point close to the solution.

Version 8.02 enhancements to PROC NLIN greatly improve its performance for these tests. In this release, PROC NLIN converges to a solution for 23 out of 27 nonlinear regression problems from a starting point far from the solution. In addition, the procedure converges to a solution for all 27 problems using a starting point close to the solution. The Version 8.02 release of the MODEL procedure, available in SAS/ETS software, produces estimates that are similar to the Version 8.02 NLIN procedure.

Appendix 1 displays the results for the nonlinear regression problems.

## Random Number Generators

The SAS System provides a number of functions for random number generation. These, too, have received attention from reviewers. The "DIEHARD Battery of Randomness Tests" (Marsaglia, 1996) provide a wide range of tests for evaluating random number generators in statistical software.   The DIEHARD program enables you to test random integers produced by the DIEHARD software itself or by another random number generator such as SAS software. To learn more about the DIEHARD battery of tests and its process for testing software that generates random numbers, download the DIEHARD software and documentation from

### http://stat.fsu.edu/~geo/.

Base SAS provides numerous random generators including the RANUNI function, which returns a random value from a uniform distribution.  The RANUNI function uses a prime modulus multiplicative congruential generator that has been well known for over 35 years.  The generator for the RANUNI function, with modulus 2^31-1, obtains the longest possible period among generators of this type that use double precision.  For any random number generator with a much larger modulus, the precision of the software would have to be greater than double precision.  In addition, another drawback of random number generators with larger moduli is that their theoretical properties are not nearly as well known.

To exercise the random number generator, the RANNUI function was used to generate 3,000,000 random numbers. These results were then tested against the DIEHARD software. The random numbers generated by the RANUNI function passed 17 of 18 DIEHARD tests for randomness, which presume a period of $2^{32}$.

For more details on the prime modulus multiplicative generator used by the RANUNI function, refer to Fishman and Moore (1982).

## Statistical Distribution Functions

Numerous functions in the DATA step enable you to calculate various quantities of statistical distributions. Release 6.12 of the SAS System introduced several new functions. The CDF function computes the cumulative distribution function for over 20 distributions, including the binomial, chi-squared, exponential, $F$, gamma, normal, $t$, uniform, and Weibull.  The SDF function computes the survival function (the upper tail of a probability distribution), and the LOGSDF computes the logarithm of the survival function, which is useful for calculating the extreme upper tails of a cumulative distribution function.  Finally, the PDF function computes

probability density (mass) functions.  The PDF, SDF, and LOGSDF functions provide access to the same distributions handled by the CDF function.

As an example of how the SDF and LOGSDF functions enable you to compute accurate probabilities, consider calculating an upper tail probability for a random variable X that is distributed as chi-squared with 100 degrees of freedom. Using the CDF function, you could compute the upper tail probability as

$$P(X > 265) = 1 - CDF('chisquared', 265, 100)$$
$$= 1.1102E{-}16,$$

but this answer is incorrect due to cancellation error.

However, if you use the SDF function,

$$P(X > 265) = SDF('chisquared', 265, 100)$$
$$= 7.2119E{-}17$$

the result is accurate to at least 10 digits of relative precision.

You can also use Base SAS statistical distribution functions to accurately compute probabilities based on extreme values of distribution parameters.  The following SAS DATA step computes probability values from the $F$-distribution with 201 numerator degrees of freedom and 10001 denominator degrees of freedom, where X decreases from 1.0 to 0.1 by 0.1.

```
data _null_;
  do x=1 to 0.1 by -0.1;
    p = probf(x,201,10001);
    put x f5.2 '  ' p e23.15;
  end;
run;
```

| X | P |
|---|---|
| 1.0 | 5.1287070378E−01 |
| 0.9 | 1.5966623195E−01 |
| 0.8 | 1.7628067051E−02 |
| 0.7 | 4.5779354283E−04 |
| 0.6 | 1.6420856256E−06 |
| 0.5 | 3.7174084640E−10 |
| 0.4 | 1.4681859564E−15 |
| 0.3 | 9.2465145773E−24 |
| 0.2 | 4.3786576061E−37 |
| 0.1 | 5.9650913649E−63 |

**Table 2:** Results Computed by the PROBF Function

Table 2 displays the results computed by the PROBF function, which are accurate to at least 10 digits of relative precision.  Note that McCullough's (1999) section on statistical distribution functions contained an inaccuracy concerning the PROBF results.  In fact, the results displayed in Table 2 are in agreement with the correct values provided in McCullough's paper.

The algorithms that compute the cumulative distribution function of the many distributions available can be very complex. In the SAS System, if a function's algorithm does not converge, the result is set to a missing value. There are two reasons why the algorithm may not converge: the intermediate numbers may be beyond the numerical limitation of the hardware, or the algorithm may have hit an internal limitation imposed by the function. For example, if the function is evaluating the probability by summing an infinite series, it sets a limit on the number of terms used in the series summation. If the series does not converge within this preset limit, the result is set to a missing value. In most of these cases, a normal approximation of the value is an appropriate alternative solution. There are many approximations available in the literature.

## Discussion

Many computing applications require a trade-off between speed and accuracy of numerical algorithms. The algorithms implemented by the analytical procedures in SAS software are optimized for performance and accuracy, efficiently producing accurate results in most situations. In rare instances, you may need to perform an analysis of variance or regression on data that are so ill-conditioned that the standard procedure like PROC ANOVA and PROC REG don't apply. In those cases, SAS users should use the ORTHOREG procedure, which appropriately handles the types of pathological data represented by the NIST Statistical Resource Datasets. In addition, developers of analytical SAS software are constantly reevaluating and improving the numerical accuracy of statistical algorithms.

Please contact Technical Support if you have additional questions.

## Appendix 1

| Test | Start | $\lambda_\beta$ | $\lambda_\sigma$ | $\lambda_r$ |
|---|---|---|---|---|
| | | PROC NLIN in SAS 6.12 Software | | |
| Misra1a | I | 9.2 | 8.9 | 10.5 |
| Chwirut2 | I | 8.3 | 8.7 | 11.2 |
| Chwirut1 | I | 7.5 | 7.9 | 11.4 |
| Lanczos3 | I | 6.7 | 6.7 | 10.6 |
| Gauss1 | I | 8.7 | 8.5 | 11.6 |
| Gauss2 | I | 8.4 | 8.1 | 10.6 |
| DanWood | I | 10 | 10.1 | 11.7 |
| Misra1b | I | 10.1 | 9.9 | 11.2 |
| Kirby2 | I | 7.4 | 7.8 | 11.6 |
| Hahn1 | I | 7.8 | 8.6 | 10.6 |
| Nelson | I | 7.1 | 7.2 | 10.9 |
| MGH17 | II | 8.8 | 8.3 | 11.5 |
| Lanczos1 | I | 10.6 | 3.3 | 3 |
| Lanczos2 | I | 10.3 | 9 | 10.2 |
| Gauss3 | I | 9.2 | 8.6 | 11 |
| Misra1c | I | 10.5 | 10.5 | 11.1 |
| Misra1d | I | 8.7 | 8.4 | 11.2 |
| Roszman1 | I | 8.6 | 9.1 | 12.2 |
| ENSO | I | 7.1 | 8.3 | 11.4 |
| MGH09 | II | 6.5 | 6.6 | 11.6 |
| Thurber | I | 6.4 | 5.8 | 11.3 |
| BoxBOD | II | 7.1 | 7.1 | 10.4 |
| Rat42 | I | 8.4 | 8 | 11.9 |
| MGH10 | II | 8.6 | 8.6 | 11.9 |
| Eckerle4 | II | 8.3 | 8.3 | 10.7 |
| Rat43 | II | 7 | 7.1 | 11.4 |
| Bennett5 | II | .4 | 0 | .8 |

| Test | Start | $\lambda_\beta$ | $\lambda_\sigma$ | $\lambda_r$ |
|---|---|---|---|---|
| | | PROC NLIN SAS 8.02 Software | | |
| Misra1a | I | 8.6 | 8.3 | 10.5 |
| Chwirut2 | I | 8.3 | 8.7 | 11.2 |
| Chwirut1 | I | 7.5 | 7.9 | 11.4 |
| Lanczos3 | I | 7.4 | 7.4 | 10.6 |
| Gauss1 | I | 8.7 | 8.5 | 11.6 |
| Gauss2 | I | 8.4 | 8.1 | 10.6 |
| DanWood | I | 10 | 10.1 | 11.7 |
| Misra1b | I | 9.4 | 9.2 | 11.3 |
| Kirby2 | I | 7.4 | 7.8 | 11.6 |
| Hahn1 | I | 7.7 | 8.5 | 10.6 |
| Nelson | I | 9.2 | 9.3 | 10.9 |
| MGH17 | II | 8.1 | 7.7 | 11.5 |
| Lanczos1 | I | 10.6 | 3.3 | 3 |
| Lanczos2 | I | 10.2 | 9.7 | 10.3 |
| Gauss3 | I | 9.2 | 8.6 | 11 |
| Misra1c | I | 9.8 | 9.7 | 11.1 |
| Misra1d | I | 8.8 | 8.5 | 11.2 |
| Roszman1 | I | 7.2 | 7.7 | 12.2 |
| ENSO | I | 6.6 | 7.7 | 11.4 |
| MGH09 | II | 6.6 | 6.7 | 11.6 |
| Thurber | I | 6.5 | 5.9 | 11.3 |
| BoxBOD | I | 6.7 | 6.7 | 10.4 |
| Rat42 | I | 8.3 | 7.9 | 11.9 |
| MGH10 | II | 10.9 | 9.6 | 13.1 |
| Eckerle4 | II | 8.3 | 8.3 | 10.7 |
| Rat43 | I | 6.8 | 6.9 | 11.4 |
| Bennett5 | I | 10.9 | 6.6 | 11.1 |

# References

Bjorck, A. (1996*), Numerical Methods for Least Squares Problems*, SIAM, Philadelphia: 1996.

Fishman, G. S. and Moore, L. R. (198.02), "A Statistical Evaluation of Multiplicative Congruential Generators with Modulus ($2^{31}-1$)," *Journal of the American Statistical Association*, 77, 129-136.

Gentleman, W. M. (1972), "Basic Procedures for Large, Sparse, or Weighted Least Squares Problems," University of Waterloo Report CSRR-2068.0, Waterloo, Ontario, Canada.

Gentleman, W. M. (1973) "Least Squares Computations by Givens Transformations with Square Roots," *Journal. Institute of Mathematics and its Applications*, 12, 329-336.

Higham, N. J., *Accuracy and Stability of Numerical Algorithms, SIAM,* Philadelphia: 1996

Longley, J. W. (1967), "An Appraisal of Least Squares Programs for the Electronic Computer from the Point of View of the User," *Journal of the American Statistical Association*, 62, 8.019-8.041.

Marsaglia, G. (1996) "DIEHARD: A Battery of Tests of Randomness," [http://stat.fsu.edu/pub/~diehard/]

McCullough, B. D. (1998.0), "Assessing the Reliability of Statistical Software, Part I," *The American Statistician*, 52, 358.0-366.

McCullough, B. D. (1999), "Assessing the Reliability of Statistical Software, Part II," *The American Statistician*, 53, 149-159.

National Institute of Standards and Technology (1997), "Statistical Reference Datasets," [http://www.nist.gov/itl/div8.098.0/strd].

Thisted, R. A. (198.08.0), *Elements of Statistical Computing: Numerical Computation*, CRC. [complete] --N

Wampler, R. H. (1970), "A Report of the Accuracy of Some Widely Used Least Squares Computer Programs," *Journal of the American Statistical Association*, 65, 549-563.

*Version 2. May 1, 2003*