UNIVERSITY OF QUEENSLAND

Prentice Computer Centre

NEWSLETTER

authorization: Director of the Centre

1 CHRISTMAS/NEW YEAR ARRANGEMENTS

No work will be processed for users on Friday, 24 December, to allow the Centre to complete end of year procedures. Users may collect work processed the previous evening up until 1 p.m. on Friday.

The Centre will close on 27 and 28 December and on 3 January. A single shift operation between 9 a.m. and 5 p.m. will be provided on 29, 30 and 31 December.

2 ALGOL V6A

Since transferring Algol V6A to STD: some problems have been encountered. For further details see HLP:NEWALG.HLP.

Our thanks to Mr. Geoff Foster of TEDI for isolating these problems and reporting them to us.

3 KNOWN ERRORS IN SPSS

The following information, obtained from SPSS Incorporated, the originators of the package, is a list of known errors currently existing in the package. Except where indicated, these should be fixed in the next version release, sometime in the first half of next year.

Though many are obscure and unlikely to arise in practice, attention is drawn to the comments on the procedures ANOVA and DISCRIMINANT. In particular, if classification function co-efficients are desired then one of the other methods (WILKS etc.) must be used in conjunction with option 12 of subprogram DISCRIMINANT.

ANOVA. Option 9 (regression approach) calculates sums of squares incorrectly and should not be used. In option 10 (hierarchical approach), all covariate effects are adjusted for all other covariates rather than only for those entered earlier in the list as indicated in the manual.

BREAKDOWN. In integer mode, both varieties of BREAKDOWN will sometimes print negative variances when the variance should be zero, and the FASTBREAK option will produce FORTRAN error messages indicating attempts to take the square roots of negative numbers.

CANCORR. All chi-square statistics are exaggerated by a factor of N/(N-1) where N is the number of cases. The number of canonical variates is limited to 5.

DATA LIST. Although the input format constructed by DATA LIST can be up to 1600 characters long, only the first 800 characters are printed.

DISCRIMINANT. A one-dimensional scatter plot, rather than a histogram, is produced by option 7 when there is only one discriminant function. Classification of data into groups does not work when the discrimination is one variable only. Classification function coefficients are not printed when METHOD=DIRECT is used. It is not known at this point whether these errors will be corrected in Release 7.0.

DO REPEAT. No check for unequal numbers of variables is made on the last list of variables. No message is produced, and the transformations included inside the DO REPEAT can generate garbage if the last list is short.

END REPEAT. A blank card following the END REPEAT will cause a spurious error message.

FREQUENCIES. The program will loop in printing histograms if any value has a frequency greater than 2 billion (which can happen in weighted files). In integer mode, the program will fail if a value range (maximum value - minimum value + 1) exceeds 32767. In general mode, it will fail if more than 32767 values are encountered for all variables. These limitations will be checked in Release 7.0 and processing will be skipped if they are violated.

LIST FILEINFO. If more than 200 value labels have been entered for a variable, the first 199 and the last variable entered are printed. In Release 7.0, this will be changed so that the last 200 are printed (since those will be saved if a SAVE FILE is done).

ONEWAY. When orthogonal polynomials are requested and the cell sizes are not equal, the sums of squares for deviation from the polynomial terms are inorrect (in fact, they can be negative). The sums of squares for the polynomial terms themselves are correct. Significance levels for range tests are limited to .001 through .20. The program can compute negative sums of squares because of precision limitations and does not protect itself very well, so it can generate FORTRAN error messages about negative square root and logarithm arguments. The program will fail if more than ten contrasts are specified.

OSIRIS VARS. Large or negative missing values are not correctly interpreted. The constructed input format is limited to 800 characters rather than 1600.

PEARSON CORR. This subprogram will fail if more than 32767 pairs are specified. The correlations produced are slightly less precise than those from REGRESSION or PARTIAL CORR because some of the intermediate calculations are done in single precision.

SAMPLE and *SAMPLE. An argument of 1 will be treated as .01.

SAVE FILE. If more than 200 value labels have been defined for a variable, the first 199 and the last one will be saved. In Release 7.0, the last 200 will be saved.

SCATTERGRAM. The program will fail if only one variable is specified in a variable list.

Transformations (IF, COMPUTE). The sequence "-0" will produce an SPSS error message indicating a missing operator.

VALUE LABELS. The keyword ALL is not accepted.

WRITE FILEINFO. The program may fail if it tries to process a DOCUMENT which originally occupied a number of cards evenly divisible by 10. A correction for this will be distributed shortly. The program also destroys any tabular alignment used in DOCUMENT cards.

General errors. If a FREQUENCIES specification does not start with INTEGER or GENERAL, or if a CROSSTABS or BREAKDOWN specification does not start with VARIABLES= or TABLES=, any OPTIONS or STATISTICS card following the procedure card will cause SPSS error 1 (unrecognized control field). The program will fail if a READ INPUT DATA card is encountered without a prior VARIABLE LIST and INPUT FORMAT card or DATA LIST.

4 IMSL versus SSP

At the present time, the Centre supports two libraries of subroutines suitable for calling from Fortran programs. These are the recently acquired International Mathematical and Statistical Library (IMSL) and the long established Scientific Subroutine Package (SSP).

In the interests of demonstrating the superiority of their routines, IMSL has distributed a series of comparative studies between routines from both libraries. We reproduce one such report here.

Linear Algebraic Equation Subroutines, Case Study #2

Let us consider one of the simplest, yet probably the most useful, numerical applicatin in existence today - that of solving linear algebraic systems. After all, everyone knows how to solve linear equations. One uses Gaussian elimination with partial pivoting followed by iterative improvement. However, even here there are state-of-the-art facts which are widely unappreciated.

Iterative improvement, for instance, is costly in both time and space. In fact it doubles the core requires and adds about 25% to the time. One often performs iterative improvement to find out that it wasn't needed. As a test, it's expensive. However, there is a much better test; it costs almost no time and space. This test determines if the answer is good enough without iterative improvement. It is used in all IMSL linear equation solvers and should be an automatic part of any good linear equation solver. This test was first reported in the JACM in January 1967 by Chartres and Geuder.

Secondly, Gaussian elimination is just a string of inner-products. In general, we know that it's a good idea to accumulate inner-products in higher precision than the rest of the computation. But, in order to accumulate double precision inner products in Gaussian elimination, one must either triple the core requirements or reorganize the computation to use a Crout algorithm. Years ago, when people computed linear equations on desk calculators, students were trained in th Crout algorithm. This training was lost with the computer and numerical analysts have had to reinitiate it.

Let's now look at test results from routines which do not implement these methods and those which do. Hilbert matrices are often used to test linear equation solvers. The condition of a Hilbert matrix increases as EXP (3.5* ORDER-OF-MATRIX). Of course one has to be careful in truncating the elements of the Hilbert matrix to obtain a matrix which is machine representable. Figure 1 plots average decimal places of accuracy against order of the matrix for SSP routines GELG and SIMQ and IMSL routine LEQT1F. The graph shows what one should expect. As the order increases, all these routines lose accuracy. Since the matrices are small, double precision accumulation of inner products sometimes does nothing, but sometimes gives more than two extra decimal digits of accuracy. For instance, in the 8 x 8 case, both SSP routines give no accuracy while the IMSL routine averaged over two places at a marginal cost in time.

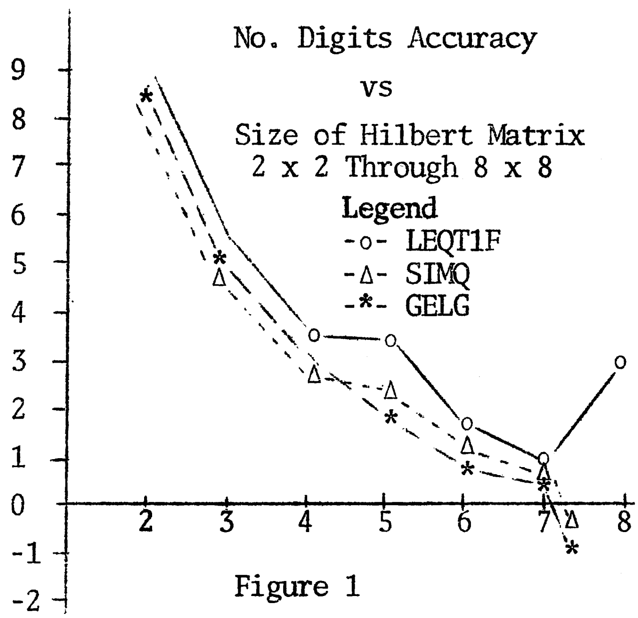

Figure 1

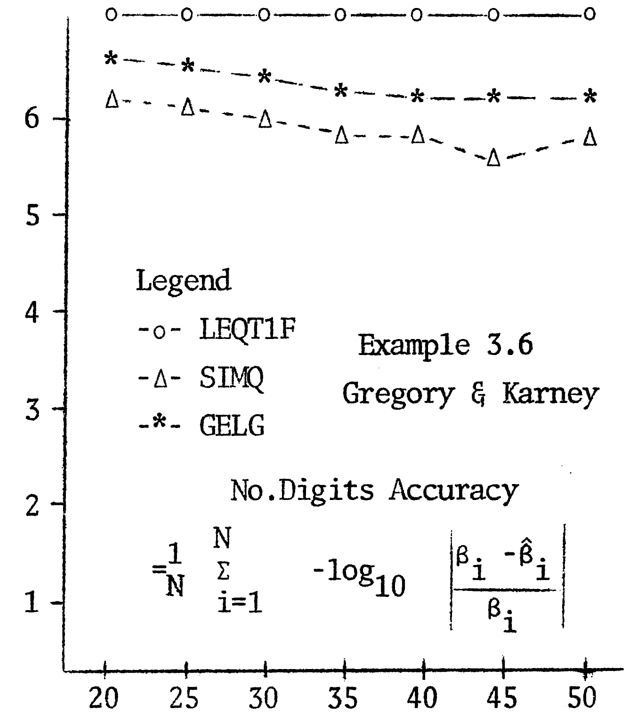

Figure 2

Figure 2 illustrates the well conditioned case for matrices of order 20 through 50. Here the IMSL routine is maintaining single precision accuracy while the accuracy of the other routines deteriorates with the increase in order. The matrices tested here are drawn from the Gregory, Karney collection.

* * * * * *

THE DIRECTOR AND STAFF OF THE PRENTICE COMPUTER CENTRE

WISH ALL USERS A MERRY CHRISTMAS AND A HAPPY NEW YEAR