NAG Fortran Library Routine Document

G03ZAF

Note: before using this routine, please read the Users' Note for your implementation to check the interpretation of ***bold italicised*** terms and other implementation-dependent details.

1 Purpose

G03ZAF produces standardized values (z-scores) for a data matrix.

2 Specification

```
SUBROUTINE G03ZAF(N, M, X, LDX, NVAR, ISX, S, E, Z, LDZ, IFAIL)
INTEGER           N, M, LDX, NVAR, ISX(M), LDZ, IFAIL
real              X(LDX,M), S(M), E(M), Z(LDZ,NVAR)
```

3 Description

For a data matrix, X, consisting of n observations on p variables, with elements x_{ij}, G03ZAF computes a matrix, Z, with elements z_{ij} such that:

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j}, \quad i = 1, 2, \ldots, n; \quad j = 1, 2, \ldots, p,$$

where μ_j is a location shift and σ_j is a scaling factor. Typically μ_j will be the mean and σ_j will be the standard deviation of the jth variable and therefore the elements in column j of Z will have zero mean and unit variance.

4 References

None.

5 Parameters

1: N – INTEGER *Input*

On entry: the number of observations in the data matrix, n.

Constraint: N $\geq$ 1.

2: M – INTEGER *Input*

On entry: the number of variables in the data array X.

Constraint: M $\geq$ NVAR.

3: X(LDX,M) – ***real*** array *Input*

On entry: X(i,j) must contain the ith sample point for the jth variable, x_{ij}, for $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots,$ M.

4: LDX – INTEGER *Input*

On entry: the first dimension of the array X as declared in the (sub)program from which G03ZAF is called.

Constraint: LDX $\geq$ N.

5: NVAR – INTEGER *Input*

On entry: the number of variables to be standardized, p.

Constraint: NVAR ≥ 1.

6: ISX(M) – INTEGER array *Input*

On entry: ISX(j) indicates whether or not the observations on the jth variable are included in the matrix of standardized values.

If ISX(j) $\neq 0$, then the observations from the jth variable are included.

If ISX(j) $= 0$, then the observations from the jth variable are not included.

Constraint: ISX(j) $\neq 0$ for NVAR values of j.

7: S(M) – ***real*** array *Input*

On entry: if ISX(j) $\neq 0$, then S(j) must contain the scaling (standard deviation), σ_j, for the jth variable.

If ISX(j) $= 0$, then S(j) is not referenced.

Constraint: if ISX(j) $\neq 0$, then S(j) > 0.0, for $j = 1, 2, \ldots, \mathrm{M}$.

8: E(M) – ***real*** array *Input*

On entry: if ISX(j) $\neq 0$, then E(j) must contain the location shift (mean), μ_j, for the jth variable.

If ISX(j) $= 0$, then E(j) is not referenced.

9: Z(LDZ,NVAR) – ***real*** array *Output*

On exit: the matrix of standardized values (z-scores), Z.

10: LDZ – INTEGER *Input*

On entry: the first dimension of the array Z as declared in the (sub)program from which G03ZAF is called.

Constraint: LDZ $\geq$ N.

11: IFAIL – INTEGER *Input/Output*

On entry: IFAIL must be set to 0, -1 or 1. Users who are unfamiliar with this parameter should refer to Chapter P01 for details.

On exit: IFAIL $= 0$ unless the routine detects an error (see Section 6).

For environments where it might be inappropriate to halt program execution when an error is detected, the value -1 or 1 is recommended. If the output of error messages is undesirable, then the value 1 is recommended. Otherwise, for users not familiar with this parameter the recommended value is 0. **When the value -1 or 1 is used it is essential to test the value of IFAIL on exit.**

6 Error Indicators and Warnings

If on entry IFAIL $= 0$ or -1, explanatory error messages are output on the current error message unit (as defined by X04AAF).

Errors or warnings detected by the routine:

IFAIL = 1

On entry, N < 1,
or NVAR < 1,
or M < NVAR,
or LDX < N,
or LDZ < N.

IFAIL = 2

On entry, there are not precisely NVAR elements of ISX ≠ 0.

IFAIL = 3

On entry, ISX(j) ≠ 0 and S(j) ≤ 0.0 for some j.

7 Accuracy

Standard accuracy is achieved.

8 Further Comments

Means and standard deviations may be obtained using G01AAF or G02BXF.

9 Example

A 4 by 3 data matrix is input along with location and scaling values. The first and third columns are scaled and the results printed.

9.1 Program Text

Note: the listing of the example program presented below uses ***bold italicised*** terms to denote precision-dependent details. Please read the Users' Note for your implementation to check the interpretation of these terms. As explained in the Essential Introduction to this manual, the results produced may not be identical for all implementations.

```
*     G03ZAF Example Program Text
*     Mark 15 Release. NAG Copyright 1991.
*     .. Parameters ..
      INTEGER          NIN, NOUT
      PARAMETER        (NIN=5,NOUT=6)
      INTEGER          NMAX, MMAX
      PARAMETER        (NMAX=4,MMAX=3)
*     .. Local Scalars ..
      INTEGER          I, IFAIL, J, M, N, NVAR
*     .. Local Arrays ..
      real             E(MMAX), S(MMAX), X(NMAX,MMAX), Z(NMAX,MMAX)
      INTEGER          ISX(MMAX)
*     .. External Subroutines ..
      EXTERNAL         G03ZAF
*     .. Executable Statements ..
      WRITE (NOUT,*) 'G03ZAF Example Program Results'
*     Skip headings in data file
      READ (NIN,*)
      READ (NIN,*) N, M, NVAR
      IF (M.LE.MMAX .AND. N.LE.NMAX) THEN
         DO 20 I = 1, N
            READ (NIN,*) (X(I,J),J=1,M)
   20    CONTINUE
         READ (NIN,*) (ISX(J),J=1,M)
         READ (NIN,*) (E(J),J=1,M)
         READ (NIN,*) (S(J),J=1,M)
         IFAIL = 0
*
         CALL G03ZAF(N,M,X,NMAX,NVAR,ISX,S,E,Z,NMAX,IFAIL)
*
         WRITE (NOUT,*)
```

```
         WRITE (NOUT,*) ' Standardized Values'
         DO 40 I = 1, N
            WRITE (NOUT,99999) (Z(I,J),J=1,NVAR)
   40    CONTINUE
      END IF
      STOP
*
99999 FORMAT (1X,9F8.3)
      END
```

9.2 Program Data

```
G03ZAF Example Program Data
4 3 2
15.0 0.0 1500.0
12.0 1.0 1000.0
18.0 2.0 1200.0
14.0 3.0  500.0
 1      0    1
14.75 0.0 1050.0
 2.50 0.0  420.3
```

9.3 Program Results

```
 G03ZAF Example Program Results

  Standardized Values
    0.100   1.071
   -1.100  -0.119
    1.300   0.357
   -0.300  -1.309
```