NAG Library Routine Document

G02BDF

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

G02BDF computes means and standard deviations of variables, sums of squares and cross-products about zero, and correlation-like coefficients for a set of data.

2 Specification

```
SUBROUTINE G02BDF(N, M, X, LDX, XBAR, STD, SSPZ, LDSSPZ, RZ, LDRZ,
1                 IFAIL)
INTEGER           N, M, LDX, LDSSPZ, LDRZ, IFAIL
double precision  X(LDX,M), XBAR(M), STD(M), SSPZ(LDSSPZ,M), RZ(LDRZ,M)
```

3 Description

The input data consists of n observations for each of m variables, given as an array

$$\left[x_{ij}\right], \qquad i = 1,2,\ldots,n (n \geq 2), \qquad j = 1,2,\ldots,m \qquad (m \geq 2),$$

where x_{ij} is the ith observation on the jth variable.

The quantities calculated are:

(a) Means:

$$\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}, \qquad j = 1,2,\ldots,m.$$

(b) Standard deviations:

$$s_j = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_{ij}-\bar{x}_j\right)^2}, \qquad j = 1,2,\ldots,m.$$

(c) Sums of squares and cross-products about zero:

$$\tilde{S}_{jk} = \sum_{i=1}^{n} x_{ij}x_{ik}, \qquad j,k = 1,2,\ldots,m.$$

(d) Correlation-like coefficients:

$$\tilde{R}_{jk} = \frac{\tilde{S}_{jk}}{\sqrt{\tilde{S}_{jj}\tilde{S}_{kk}}}, \qquad j,k = 1,2,\ldots,m.$$

If $\tilde{S}_{jj}$ or $\tilde{S}_{kk}$ is zero, $\tilde{R}_{jk}$ is set to zero.

4 References

None.

5 Parameters

1: N – INTEGER *Input*

On entry: n, the number of observations or cases.

Constraint: N ≥ 2.

2: M – INTEGER *Input*

On entry: m, the number of variables.

Constraint: M ≥ 2.

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

On entry: X(i,j) must be set to the value of x_{ij}, the ith observation on the jth variable, for $i = 1, 2, \ldots n$ and $j = 1, 2, \ldots m$.

4: LDX – INTEGER *Input*

On entry: the first dimension of the array X as declared in the (sub)program from which G02BDF is called.

Constraint: LDX $\geq$ N.

5: XBAR(M) – ***double precision*** array *Output*

On exit: XBAR(j) contains the mean value, $\bar{x}_j$, of the jth variable, for $j = 1, 2, \ldots, m$.

6: STD(M) – ***double precision*** array *Output*

On exit: the standard deviation, s_j, of the jth variable, for $j = 1, 2, \ldots, m$.

7: SSPZ(LDSSPZ,M) – ***double precision*** array *Output*

On exit: SSPZ(j,k) is the cross-product about zero, $\tilde{S}_{jk}$, for $j, k = 1, 2, \ldots, m$.

8: LDSSPZ – INTEGER *Input*

On entry: the first dimension of the array SSPZ as declared in the (sub)program from which G02BDF is called.

Constraint: LDSSPZ $\geq$ M.

9: RZ(LDRZ,M) – ***double precision*** array *Output*

On exit: RZ(j,k) is the correlation-like coefficient, $\tilde{R}_{jk}$, between the jth and kth variables, for $j, k = 1, 2, \ldots, m$.

10: LDRZ – INTEGER *Input*

On entry: the first dimension of the array RZ as declared in the (sub)program from which G02BDF is called.

Constraint: LDRZ $\geq$ M.

11: IFAIL – INTEGER *Input/Output*

On entry: IFAIL must be set to 0, -1 or 1. If you are unfamiliar with this parameter you should refer to Section 3.3 in the Essential Introduction 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, if you are 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 < 2.

IFAIL = 2

On entry, M < 2.

IFAIL = 3

On entry, LDX < N,
or LDSSPZ < M,
or LDRZ < M.

7 Accuracy

G02BDF does not use ***additional precision*** arithmetic for the accumulation of scalar products, so there may be a loss of significant figures for large n.

8 Further Comments

The time taken by G02BDF depends on n and m.

The routine uses a two-pass algorithm.

9 Example

This example reads in a set of data consisting of five observations on each of three variables. The means, standard deviations, sums of squares and cross-products about zero, and correlation-like coefficients for all three variables are then calculated and printed.

9.1 Program Text

```
*     G02BDF Example Program Text
*     Mark 14 Revised. NAG Copyright 1989.
*     .. Parameters ..
      INTEGER          M, N, LDX, LDSSPZ, LDRZ
      PARAMETER        (M=3,N=5,LDX=N,LDSSPZ=M,LDRZ=M)
      INTEGER          NIN, NOUT
      PARAMETER        (NIN=5,NOUT=6)
*     .. Local Scalars ..
      INTEGER          I, IFAIL, J
*     .. Local Arrays ..
      DOUBLE PRECISION RZ(LDRZ,M), SSPZ(LDSSPZ,M), STD(M), X(LDX,M),
     +                 XBAR(M)
*     .. External Subroutines ..
      EXTERNAL         G02BDF
*     .. Executable Statements ..
      WRITE (NOUT,*) 'G02BDF Example Program Results'
*     Skip heading in data file
      READ (NIN,*)
      READ (NIN,*) ((X(I,J),J=1,M),I=1,N)
      WRITE (NOUT,*)
      WRITE (NOUT,99999) 'Number of variables (columns) =', M
```

```
      WRITE (NOUT,99999) 'Number of cases     (rows)    =', N
      WRITE (NOUT,*)
      WRITE (NOUT,*) 'Data matrix is:-'
      WRITE (NOUT,*)
      WRITE (NOUT,99998) (J,J=1,M)
      WRITE (NOUT,99997) (I,(X(I,J),J=1,M),I=1,N)
      WRITE (NOUT,*)
      IFAIL = 1
*
      CALL G02BDF(N,M,X,LDX,XBAR,STD,SSPZ,LDSSPZ,RZ,LDRZ,IFAIL)
*
      IF (IFAIL.LT.0) THEN
         WRITE (NOUT,*)
         WRITE (NOUT,99999) ' ** G02BDF returned with IFAIL = ', IFAIL
      ELSE
         IF (IFAIL.NE.0) THEN
            WRITE (NOUT,99999) 'Routine fails, IFAIL =', IFAIL
         ELSE
            WRITE (NOUT,*) 'Variable   Mean    St. dev.'
            WRITE (NOUT,99996) (I,XBAR(I),STD(I),I=1,M)
            WRITE (NOUT,*)
            WRITE (NOUT,*) 'Sums of squares and cross-products about'//
     +        ' zero'
            WRITE (NOUT,99998) (I,I=1,M)
            WRITE (NOUT,99997) (I,(SSPZ(I,J),J=1,M),I=1,M)
            WRITE (NOUT,*)
            WRITE (NOUT,*) 'Correlation-like coefficients'
            WRITE (NOUT,99998) (I,I=1,M)
            WRITE (NOUT,99997) (I,(RZ(I,J),J=1,M),I=1,M)
         END IF
      END IF
*
99999 FORMAT (1X,A,I5)
99998 FORMAT (1X,6I12)
99997 FORMAT (1X,I3,3F12.4)
99996 FORMAT (1X,I5,2F11.4)
      END
```

9.2 Program Data

```
G02BDF Example Program Data
2.00      3.00      3.00
4.00      6.00      4.00
9.00      9.00      0.00
0.00      12.00     2.00
12.00     -1.00     5.00
```

9.3 Program Results

```
 G02BDF Example Program Results

 Number of variables (columns) =    3
 Number of cases     (rows)    =    5

 Data matrix is:-

            1           2           3
   1      2.0000      3.0000      3.0000
   2      4.0000      6.0000      4.0000
   3      9.0000      9.0000      0.0000
   4      0.0000     12.0000      2.0000
   5     12.0000     -1.0000      5.0000

 Variable   Mean    St. dev.
     1     5.4000     4.9800
     2     5.8000     5.0695
     3     2.8000     1.9235

 Sums of squares and cross-products about zero
            1           2           3
```

```
   1    245.0000     99.0000     82.0000
   2     99.0000    271.0000     52.0000
   3     82.0000     52.0000     54.0000

 Correlation-like coefficients
           1           2           3
   1     1.0000      0.3842      0.7129
   2     0.3842      1.0000      0.4299
   3     0.7129      0.4299      1.0000
```