

# 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,      &
                   IFAIL)
INTEGER                N, M, LDX, LDSSPZ, LDRZ, IFAIL
REAL (KIND=nag_wp)    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

$$[x_{ij}], \quad i = 1, 2, \dots, n (n \geq 2), \quad j = 1, 2, \dots, m \quad (m \geq 2),$$

where  $x_{ij}$  is the  $i$ th observation on the  $j$ th variable.

The quantities calculated are:

(a) Means:

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

(b) Standard deviations:

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

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

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

(d) Correlation-like coefficients:

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

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

### 4 References

None.

## 5 Arguments

- 1: N – INTEGER *Input*  
*On entry:*  $n$ , the number of observations or cases.  
*Constraint:*  $N \geq 2$ .
  
- 2: M – INTEGER *Input*  
*On entry:*  $m$ , the number of variables.  
*Constraint:*  $M \geq 2$ .
  
- 3: X(LDX, M) – REAL (KIND=nag\_wp) array *Input*  
*On entry:*  $X(i, j)$  must be set to the value of  $x_{ij}$ , the  $i$ th observation on the  $j$ th variable, for  $i = 1, 2, \dots, n$  and  $j = 1, 2, \dots, m$ .
  
- 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) – REAL (KIND=nag\_wp) array *Output*  
*On exit:* XBAR( $j$ ) contains the mean value,  $\bar{x}_j$ , of the  $j$ th variable, for  $j = 1, 2, \dots, m$ .
  
- 6: STD(M) – REAL (KIND=nag\_wp) array *Output*  
*On exit:* the standard deviation,  $s_j$ , of the  $j$ th variable, for  $j = 1, 2, \dots, m$ .
  
- 7: SSPZ(LDSSPZ, M) – REAL (KIND=nag\_wp) array *Output*  
*On exit:* SSPZ( $j, k$ ) is the cross-product about zero,  $\tilde{S}_{jk}$ , for  $j = 1, 2, \dots, m$  and  $k = 1, 2, \dots, m$ .
  
- 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) – REAL (KIND=nag\_wp) array *Output*  
*On exit:* RZ( $j, k$ ) is the correlation-like coefficient,  $\tilde{R}_{jk}$ , between the  $j$ th and  $k$ th variables, for  $j = 1, 2, \dots, m$  and  $k = 1, 2, \dots, m$ .
  
- 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 argument you should refer to Section 3.4 in How to Use the NAG Library and its Documentation for details.  
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 argument, the

recommended value is 0. **When the value  $-1$  or  $1$  is used it is essential to test the value of IFAIL on exit.**

*On exit:* IFAIL = 0 unless the routine detects an error or a warning has been flagged (see Section 6).

## 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  $LDRZ < M$ .

IFAIL =  $-99$

An unexpected error has been triggered by this routine. Please contact NAG.

See Section 3.9 in How to Use the NAG Library and its Documentation for further information.

IFAIL =  $-399$

Your licence key may have expired or may not have been installed correctly.

See Section 3.8 in How to Use the NAG Library and its Documentation for further information.

IFAIL =  $-999$

Dynamic memory allocation failed.

See Section 3.7 in How to Use the NAG Library and its Documentation for further information.

## 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 Parallelism and Performance

G02BDF is threaded by NAG for parallel execution in multithreaded implementations of the NAG Library.

Please consult the X06 Chapter Introduction for information on how to control and interrogate the OpenMP environment used within this routine. Please also consult the Users' Note for your implementation for any additional implementation-specific information.

## 9 Further Comments

The time taken by G02BDF depends on  $n$  and  $m$ .

The routine uses a two-pass algorithm.

## 10 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.

### 10.1 Program Text

Program g02bdfc

```
!      G02BDF Example Program Text

!      Mark 26 Release. NAG Copyright 2016.

!      .. Use Statements ..
      Use nag_library, Only: g02bdf, nag_wp
!      .. Implicit None Statement ..
      Implicit None
!      .. Parameters ..
      Integer, Parameter          :: nin = 5, nout = 6
!      .. Local Scalars ..
      Integer                     :: i, ifail, ldrz, ldsspz, ldx, m, n
!      .. Local Arrays ..
      Real (Kind=nag_wp), Allocatable :: rz(:,,:), sspz(:,,:), std(:), x(:,,:), &
                                         xbar(:)

!      .. Executable Statements ..
      Write (nout,*) 'G02BDF Example Program Results'
      Write (nout,*)

!      Skip heading in data file
      Read (nin,*)

!      Read in the problem size
      Read (nin,*) n, m

      ldrz = m
      ldsspz = m
      ldx = n
      Allocate (rz(ldrz,m),sspz(ldsspz,m),std(m),x(ldx,m),xbar(m))

!      Read in data
      Read (nin,*)(x(i,1:m),i=1,n)

!      Display data
      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)(i,i=1,m)
      Write (nout,99997)(i,x(i,1:m),i=1,n)
      Write (nout,*)

!      Compute summary statistics
      ifail = 0
      Call g02bdf(n,m,x,ldx,xbar,std,sspz,ldsspz,rz,ldrz,ifail)

!      Display results
      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,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,1:m),i=1,m)
```

```

99999 Format (1X,A,I5)
99998 Format (1X,6I12)
99997 Format (1X,I3,3F12.4)
99996 Format (1X,I5,2F11.4)
      End Program g02bdfc

```

## 10.2 Program Data

G02BDF Example Program Data

```

5  3      :: N, M
2.0    3.0    3.0
4.0    6.0    4.0
9.0    9.0    0.0
0.0   12.0    2.0
12.0   -1.0    5.0      :: End of X

```

## 10.3 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

---