

# NAG Fortran Library Routine Document

## G02BGF

**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

G02BGF computes means and standard deviations, sums of squares and cross-products of deviations from means, and Pearson product-moment correlation coefficients for selected variables.

### 2 Specification

```

SUBROUTINE G02BGF(N, M, X, IX, NVAR, KVAR, XBAR, STD, SSP, ISSP, R, IR,
1                IFAIL)
  INTEGER          N, M, IX, NVARS, KVAR(NVAR), ISSP, IR, IFAIL
  real           X(IX,M), XBAR(NVAR), STD(NVAR), SSP(ISSP,NVAR),
1                R(IR,NVAR)

```

### 3 Description

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

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

where  $x_{ij}$  is the  $i$ th observation on the  $j$ th variable, together with the subset of these variables,  $v_1, v_2, \dots, v_p$ , for which information is required.

The quantities calculated are:

(a) Means:

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}, \quad j = v_1, v_2, \dots, v_p.$$

(b) Standard deviations:

$$s_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}, \quad j = v_1, v_2, \dots, v_p.$$

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

$$S_{jk} = \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k), \quad j, k = v_1, v_2, \dots, v_p.$$

(d) Pearson product-moment correlation coefficients:

$$R_{jk} = \frac{S_{jk}}{\sqrt{S_{jj}S_{kk}}}, \quad j, k = v_1, v_2, \dots, v_p.$$

If  $S_{jj}$  or  $S_{kk}$  is zero,  $R_{jk}$  is set to zero.

### 4 References

None.

## 5 Parameters

- 1: N – INTEGER *Input*  
*On entry:* the number,  $n$ , of observations or cases.  
*Constraint:*  $N \geq 2$ .
- 2: M – INTEGER *Input*  
*On entry:* the number,  $m$ , of variables.  
*Constraint:*  $M \geq 2$ .
- 3: X(IX,M) – *real* array *Input*  
*On entry:* X( $i, j$ ) must be set to  $x_{ij}$ , the value of the  $i$ th observation on the  $j$ th variable, for  $i = 1, 2, \dots, n$ ;  $j = 1, 2, \dots, m$ .
- 4: IX – INTEGER *Input*  
*On entry:* the first dimension of the array X as declared in the (sub)program from which G02BGF is called.  
*Constraint:*  $IX \geq N$ .
- 5: NVAR – INTEGER *Input*  
*On entry:* the number,  $p$ , of variables for which information is required.  
*Constraint:*  $2 \leq \text{NVAR} \leq M$ .
- 6: KVAR(NVAR) – INTEGER array *Input*  
*On entry:* KVAR( $j$ ) must be set to the column number in X of the  $j$ th variable for which information is required, for  $j = 1, 2, \dots, p$ .  
*Constraint:*  $1 \leq \text{KVAR}(j) \leq M$ , for  $j = 1, 2, \dots, p$ .
- 7: XBAR(NVAR) – *real* array *Output*  
*On exit:* the mean value,  $\bar{x}_j$ , of the variable specified in KVAR( $j$ ), for  $j = 1, 2, \dots, p$ .
- 8: STD(NVAR) – *real* array *Output*  
*On exit:* the standard deviation,  $s_j$ , of the variable specified in KVAR( $j$ ), for  $j = 1, 2, \dots, p$ .
- 9: SSP(ISSP,NVAR) – *real* array *Output*  
*On exit:* SSP( $j, k$ ) is the cross-product of deviations,  $S_{jk}$ , for the variables specified in KVAR( $j$ ) and KVAR( $k$ ), for  $j, k = 1, 2, \dots, p$ .
- 10: ISSP – INTEGER *Input*  
*On entry:* the first dimension of the array SSP as declared in the (sub)program from which G02BGF is called.  
*Constraint:*  $\text{ISSP} \geq \text{NVAR}$ .
- 11: R(IR,NVAR) – *real* array *Output*  
*On exit:* R( $j, k$ ) is the product-moment correlation coefficient,  $R_{jk}$ , between the variables specified in KVAR( $j$ ) and KVAR( $k$ ), for  $j, k = 1, 2, \dots, p$ .

12: IR – INTEGER

*Input*

*On entry:* the first dimension of the array R as declared in the (sub)program from which G02BGF is called.

*Constraint:*  $IR \geq N\text{VARS}$ .

13: 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,  $N < 2$ .

IFAIL = 2

On entry,  $N\text{VARS} < 2$ ,  
or  $N\text{VARS} > M$ .

IFAIL = 3

On entry,  $IX < N$ ,  
or  $ISSP < N\text{VARS}$ ,  
or  $IR < N\text{VARS}$ .

IFAIL = 4

On entry,  $K\text{VAR}(j) < 1$ ,  
or  $K\text{VAR}(j) > M$  for some  $j = 1, 2, \dots, N\text{VARS}$ .

## 7 Accuracy

The routine 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 the routine depends on  $n$  and  $p$ .

The routine uses a two pass algorithm.

## 9 Example

The following program reads in a set of data consisting of five observations on each of four variables. The means, standard deviations, sums of squares and cross-products of deviations from means, and Pearson product-moment correlation coefficients for the fourth, first and second variables are then calculated and 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.

```

*      G02BGF Example Program Text
*      Mark 14 Revised.  NAG Copyright 1989.
*      .. Parameters ..
INTEGER          M, N, NV, IA, ISSP, ICORR
PARAMETER       (M=4,N=5,NV=3,IA=N,ISSP=NV,ICORR=NV)
INTEGER          NIN, NOUT
PARAMETER       (NIN=5,NOUT=6)
*      .. Local Scalars ..
INTEGER          I, IFAIL, J
*      .. Local Arrays ..
real           A(IA,M), AMEAN(NV), CORR(ICORR,NV), SSP(ISSP,NV),
+              STD(NV)
INTEGER          KVAR(NV)
*      .. External Subroutines ..
EXTERNAL        G02BGF
*      .. Executable Statements ..
WRITE (NOUT,*) 'G02BGF Example Program Results'
*      Skip heading in data file
READ (NIN,*)
KVAR(1) = 4
KVAR(2) = 1
KVAR(3) = 2
READ (NIN,*) ((A(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,(A(I,J),J=1,M),I=1,N)
WRITE (NOUT,*)
IFAIL = 1
*
CALL G02BGF(N,M,A,IA,NV,KVAR,AMEAN,STD,SSP,ISSP,CORR,ICORR,IFAIL)
*
IF (IFAIL.NE.0) THEN
  WRITE (NOUT,99999) 'Routine fails, IFAIL =', IFAIL
ELSE
  WRITE (NOUT,*) 'Variable   Mean   St. dev.'
  WRITE (NOUT,99995) (KVAR(I),AMEAN(I),STD(I),I=1,NV)
  WRITE (NOUT,*)
  WRITE (NOUT,*)
+  'Sums of squares and cross-products of deviations'
  WRITE (NOUT,99998) (KVAR(I),I=1,NV)
  WRITE (NOUT,99996) (KVAR(I),(SSP(I,J),J=1,NV),I=1,NV)
  WRITE (NOUT,*)
  WRITE (NOUT,*) 'Correlation coefficients'
  WRITE (NOUT,99998) (KVAR(I),I=1,NV)
  WRITE (NOUT,99996) (KVAR(I),(CORR(I,J),J=1,NV),I=1,NV)
END IF
STOP
*
99999 FORMAT (1X,A,I2)
99998 FORMAT (1X,6I12)
99997 FORMAT (1X,I3,4F12.4)
99996 FORMAT (1X,I3,3F12.4)
99995 FORMAT (1X,I5,2F11.4)
END

```

## 9.2 Program Data

G02BGF Example Program Data

3.00	3.00	1.00	2.00
6.00	4.00	-1.00	4.00
9.00	0.00	5.00	9.00
12.00	2.00	0.00	0.00
-1.00	5.00	4.00	12.00

## 9.3 Program Results

G02BGF Example Program Results

Number of variables (columns) = 4  
 Number of cases (rows) = 5

Data matrix is:-

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

Variable	Mean	St. dev.
4	5.4000	4.9800
1	5.8000	5.0695
2	2.8000	1.9235

Sums of squares and cross-products of deviations

	4	1	2
4	99.2000	-57.6000	6.4000
1	-57.6000	102.8000	-29.2000
2	6.4000	-29.2000	14.8000

Correlation coefficients

	4	1	2
4	1.0000	-0.5704	0.1670
1	-0.5704	1.0000	-0.7486
2	0.1670	-0.7486	1.0000

---