

G02BKF – NAG Fortran Library Routine Document

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

G02BKF computes means and standard deviations, sums of squares and cross-products about zero, and correlation-like coefficients for selected variables.

2 Specification

```

SUBROUTINE G02BKF(N, M, X, IX, NVAR, KVAR, XBAR, STD, SSPZ,
1             ISSPZ, RZ, IRZ, IFAIL)
  INTEGER      N, M, IX, NVAR, KVAR(NVAR), ISSPZ, IRZ, IFAIL
  real        X(IX,M), XBAR(NVAR), STD(NVAR),
1             SSPZ(ISSPZ,NVAR), RZ(IRZ,NVAR)

```

3 Description

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

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

$$j = 1, 2, \dots, m \quad (m \geq 2).$$

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:

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

- (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 about zero:

$$\tilde{S}_{jk} = \sum_{i=1}^n x_{ij} x_{ik}, \quad j, k = v_1, v_2, \dots, v_p$$

- (d) Correlation-like coefficients:

$$\tilde{R}_{jk} = \frac{\tilde{S}_{jk}}{\sqrt{\tilde{S}_{jj} \tilde{S}_{kk}}}, \quad j, k = v_1, v_2, \dots, v_p$$

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: 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 G02BKF 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:** SSPZ(ISSPZ,NVAR) — *real* array *Output*
On exit: SSPZ(j, k) is the cross-product about zero, \tilde{S}_{jk} , for the variables specified in KVAR(j) and KVAR(k), for $j, k = 1, 2, \dots, p$.
- 10:** ISSPZ — INTEGER *Input*
On entry: the first dimension of the array SSPZ as declared in the (sub)program from which G02BKF is called.
Constraint: $\text{ISSPZ} \geq \text{NVAR}$.
- 11:** RZ(IRZ,NVAR) — *real* array *Output*
On exit: RZ(j, k) is the correlation-like coefficient, \tilde{R}_{jk} , between the variables specified in KVAR(j) and KVAR(k), for $j, k = 1, 2, \dots, p$.
- 12:** IRZ — INTEGER *Input*
On entry: the first dimension of the array RZ as declared in the (sub)program from which G02BKF is called.
Constraint: $\text{IRZ} \geq \text{NVAR}$.

13: IFAIL — INTEGER*Input/Output*

On entry: IFAIL must be set to 0, -1 or 1. For users not familiar with this parameter (described in Chapter P01) the recommended value is 0.

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

6 Error Indicators and Warnings

Errors detected by the routine:

IFAIL = 1

On entry, $N < 2$.

IFAIL = 2

On entry, $NVARS < 2$,
or $NVARS > M$.

IFAIL = 3

On entry, $IX < N$,
or $ISSPZ < NVARS$,
or $IRZ < NVARS$.

IFAIL = 4

On entry, $KVAR(j) < 1$,
or $KVAR(j) > M$ for some $j = 1, 2, \dots, NVARS$.

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 about zero, and correlation-like 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.

```
*      G02BKF 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       G02BKF
*   .. Executable Statements ..
WRITE (NOUT,*) 'G02BKF Example Program Results'
*   Skip heading in data file
READ (NIN,*)
READ (NIN,*) ((A(I,J),J=1,M),I=1,N)
KVAR(1) = 4
KVAR(2) = 1
KVAR(3) = 2
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 G02BKF(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 about zero'
  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-like 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,I3)
99998 FORMAT (1X,4I12)
99997 FORMAT (1X,I3,4F12.4)
99996 FORMAT (1X,I3,3F12.4)
99995 FORMAT (1X,I5,2F11.4)
END

```

9.2 Program Data

G02BKF 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

G02BKF 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 about zero

	4	1	2
4	245.0000	99.0000	82.0000
1	99.0000	271.0000	52.0000
2	82.0000	52.0000	54.0000

Correlation-like coefficients

	4	1	2
4	1.0000	0.3842	0.7129
1	0.3842	1.0000	0.4299
2	0.7129	0.4299	1.0000