NAG Library Routine Document

G02BKF

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, LDX, NVARS, KVAR, XBAR, STD, SSPZ, LDSSPZ,     &
                   RZ, LDRZ, IFAIL)

INTEGER             N, M, LDX, NVARS, KVAR(NVARS), LDSSPZ, LDRZ, IFAIL
REAL (KIND=nag_wp) X(LDX,M), XBAR(NVARS), STD(NVARS),                       &
                   SSPZ(LDSSPZ,NVARS), RZ(LDRZ,NVARS)
```

3 Description

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

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

where x_{ij} is the ith observation on the jth variable, together with the subset of these variables, $v_1, v_2, \ldots, v_p$, for which information is required.

The quantities calculated are:

(a) Means:

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

(b) Standard deviations:

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

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

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

(d) Correlation-like coefficients:

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

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 ≥ 2.

2: M – INTEGER *Input*

On entry: m, the number of variables.

Constraint: M ≥ 2.

3: X(LDX, M) – REAL (KIND=nag_wp) array *Input*

On entry: X(i,j) must be set to x_{ij}, the value of 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 G02BKF is called.

Constraint: LDX $\geq$ N.

5: NVARS – INTEGER *Input*

On entry: p, the number of variables for which information is required.

Constraint: $2 \leq$ NVARS $\leq$ M.

6: KVAR(NVARS) – INTEGER array *Input*

On entry: KVAR(j) must be set to the column number in X of the jth variable for which information is required, for $j = 1,2,\ldots,p$.

Constraint: $1 \leq$ KVAR$(j) \leq$ M, for $j = 1,2,\ldots,p$.

7: XBAR(NVARS) – REAL (KIND=nag_wp) array *Output*

On exit: the mean value, $\bar{x}_j$, of the variable specified in KVAR(j), for $j = 1,2,\ldots,p$.

8: STD(NVARS) – REAL (KIND=nag_wp) array *Output*

On exit: the standard deviation, s_j, of the variable specified in KVAR(j), for $j = 1,2,\ldots,p$.

9: SSPZ(LDSSPZ, NVARS) – REAL (KIND=nag_wp) 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 $j = 1,2,\ldots,p$ and $k = 1,2,\ldots,p$.

10: LDSSPZ – INTEGER *Input*

On entry: the first dimension of the array SSPZ as declared in the (sub)program from which G02BKF is called.

Constraint: LDSSPZ $\geq$ NVARS.

11: RZ(LDRZ, NVARS) – REAL (KIND=nag_wp) 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 $j = 1,2,\ldots,p$ and $k = 1,2,\ldots,p$.

12: LDRZ – INTEGER *Input*

On entry: the first dimension of the array RZ as declared in the (sub)program from which G02BKF is called.

Constraint: LDRZ $\geq$ NVARS.

13: 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, NVARS < 2,
or NVARS $>$ M.

IFAIL $= 3$

On entry, LDX $<$ N,
or LDSSPZ $<$ NVARS,
or LDRZ $<$ NVARS.

IFAIL $= 4$

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

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

G02BKF 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

G02BKF is not threaded in any implementation.

9 Further Comments

The time taken by G02BKF depends on n and p.

The routine uses a two-pass algorithm.

10 Example

This example 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.

10.1 Program Text

```
    Program g02bkfe

!     G02BKF Example Program Text

!     Mark 26 Release. NAG Copyright 2016.

!     .. Use Statements ..
      Use nag_library, Only: g02bkf, nag_wp
!     .. Implicit None Statement ..
      Implicit None
!     .. Parameters ..
      Integer, Parameter               :: nin = 5, nout = 6
!     .. Local Scalars ..
      Integer                          :: i, ifail, ldrz, ldsspz, ldx, m, n,   &
                                          nvars
!     .. Local Arrays ..
      Real (Kind=nag_wp), Allocatable  :: rz(:,:), sspz(:,:), std(:), x(:,:),  &
                                          xbar(:)
      Integer, Allocatable             :: kvar(:)
!     .. Executable Statements ..
      Write (nout,*) 'G02BKF Example Program Results'
      Write (nout,*)

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

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

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

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

!     Read in column IDs
      Read (nin,*) kvar(1:nvars)

!     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,*)

!     Calculate summary statistics
      ifail = 0
      Call g02bkf(n,m,x,ldx,nvars,kvar,xbar,std,sspz,ldsspz,rz,ldrz,ifail)

!     Display results
      Write (nout,*) 'Variable   Mean    St. dev.'
      Write (nout,99995)(kvar(i),xbar(i),std(i),i=1,nvars)
      Write (nout,*)
      Write (nout,*) 'Sums of squares and cross-products about' // ' zero'
      Write (nout,99998) kvar(1:nvars)
      Write (nout,99996)(kvar(i),sspz(i,1:nvars),i=1,nvars)
      Write (nout,*)
      Write (nout,*) 'Correlation-like coefficients'
      Write (nout,99998) kvar(1:nvars)
      Write (nout,99996)(kvar(i),rz(i,1:nvars),i=1,nvars)

99999 Format (1X,A,I5)
99998 Format (1X,4I12)
99997 Format (1X,I3,4F12.4)
99996 Format (1X,I3,3F12.4)
99995 Format (1X,I5,2F11.4)
    End Program g02bkfe
```

10.2 Program Data

```
G02BKF Example Program Data
5  4  3                    :: N, M, NVARS
 3.0    3.0    1.0    2.0
 6.0    4.0   -1.0    4.0
 9.0    0.0    5.0    9.0
12.0    2.0    0.0    0.0
-1.0    5.0    4.0   12.0  :: End of X
4  1  2                    :: KVAR
```

10.3 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
```