NAG Library Routine Document

G02BAF

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

G02BAF computes means and standard deviations of variables, sums of squares and cross-products of deviations from means, and Pearson product-moment correlation coefficients for a set of data.

2 Specification

```
SUBROUTINE G02BAF (N, M, X, LDX, XBAR, STD, SSP, LDSSP, R, LDR, IFAIL)
INTEGER            N, M, LDX, LDSSP, LDR, IFAIL
REAL (KIND=nag_wp) X(LDX,M), XBAR(M), STD(M), SSP(LDSSP,M), R(LDR,M)
```

3 Description

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

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

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

The quantities calculated are:

(a) Means:

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

(b) Standard deviations:

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

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

$$S_{jk} = \sum_{i=1}^{n}\big(x_{ij} - \bar{x}_j\big)\big(x_{ik} - \bar{x}_k\big), \quad j,k = 1,2,\ldots,m.$$

(d) Pearson product-moment correlation coefficients:

$$R_{jk} = \frac{S_{jk}}{\sqrt{S_{jj}S_{kk}}}, \quad j,k = 1,2,\ldots,m.$$

If S_{jj} or S_{kk} is zero, 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 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 G02BAF is called.

Constraint: LDX $\geq$ N.

5: XBAR(M) – REAL (KIND=nag_wp) array *Output*

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

6: STD(M) – REAL (KIND=nag_wp) array *Output*

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

7: SSP(LDSSP, M) – REAL (KIND=nag_wp) array *Output*

On exit: SSP(j,k) is the cross-product of deviations S_{jk}, for $j = 1, 2, \ldots, m$ and $k = 1, 2, \ldots, m$.

8: LDSSP – INTEGER *Input*

On entry: the first dimension of the array SSP as declared in the (sub)program from which G02BAF is called.

Constraint: LDSSP $\geq$ M.

9: R(LDR, M) – REAL (KIND=nag_wp) array *Output*

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

10: LDR – INTEGER *Input*

On entry: the first dimension of the array R as declared in the (sub)program from which G02BAF is called.

Constraint: LDR $\geq$ 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 LDSSP $<$ M,
or LDR $<$ 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

G02BAF 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

G02BAF 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 G02BAF 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 of deviations from means, and Pearson product-moment correlation coefficients for all three variables are then calculated and printed.

10.1 Program Text

```
    Program g02bafe

!     G02BAF Example Program Text

!     Mark 26 Release. NAG Copyright 2016.

!     .. Use Statements ..
      Use nag_library, Only: g02baf, nag_wp
!     .. Implicit None Statement ..
      Implicit None
!     .. Parameters ..
      Integer, Parameter               :: nin = 5, nout = 6
!     .. Local Scalars ..
      Integer                          :: i, ifail, ldr, ldssp, ldx, m, n
!     .. Local Arrays ..
      Real (Kind=nag_wp), Allocatable  :: r(:,:), ssp(:,:), std(:), x(:,:),    &
                                          xbar(:)
!     .. Executable Statements ..
      Write (nout,*) 'G02BAF Example Program Results'
      Write (nout,*)

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

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

      ldr = m
      ldssp = m
      ldx = n
      Allocate (x(ldx,m),r(ldr,m),ssp(ldssp,m),std(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 g02baf(n,m,x,ldx,xbar,std,ssp,ldssp,r,ldr,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 of deviations'
      Write (nout,99998)(i,i=1,m)
      Write (nout,99997)(i,ssp(i,1:m),i=1,m)
      Write (nout,*)
      Write (nout,*) 'Correlation coefficients'
      Write (nout,99998)(i,i=1,m)
      Write (nout,99997)(i,r(i,1:m),i=1,m)
```

```
99999 Format (1X,A,IO)
99998 Format (1X,6I12)
99997 Format (1X,I3,3F12.4)
99996 Format (1X,I5,2F11.4)
    End Program g02bafe
```

10.2 Program Data

```
G02BAF 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

```
 G02BAF 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 of deviations
            1           2           3
   1     99.2000    -57.6000      6.4000
   2    -57.6000    102.8000    -29.2000
   3      6.4000    -29.2000     14.8000

 Correlation coefficients
            1           2           3
   1      1.0000     -0.5704      0.1670
   2     -0.5704      1.0000     -0.7486
   3      0.1670     -0.7486      1.0000
```