

NAG Fortran 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, IX, XBAR, STD, SSP, ISSP, R, IR, IFAIL)
INTEGER          N, M, IX, ISSP, IR, IFAIL
real            X(IX,M), XBAR(M), STD(M), SSP(ISSP,M), R(IR,M)
```

3 Description

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

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

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 of deviations from means:

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

(d) Pearson product-moment correlation coefficients:

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

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 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 G02BAF is called.
Constraint: $IX \geq N$.
- 5: XBAR(M) – *real* array *Output*
On exit: the mean value, $\bar{x}_j$, of the j th variable, for $j = 1, 2, \dots, m$.
- 6: STD(M) – *real* array *Output*
On exit: the standard deviation, s_j , of the j th variable, for $j = 1, 2, \dots, m$.
- 7: SSP(ISSP,M) – *real* array *Output*
On exit: $SSP(j, k)$ is the cross-product of deviations S_{jk} , for $j, k = 1, 2, \dots, m$.
- 8: ISSP – INTEGER *Input*
On entry: the first dimension of the array SSP as declared in the (sub)program from which G02BAF is called.
Constraint: $ISSP \geq M$.
- 9: R(IR,M) – *real* array *Output*
On exit: $R(j, k)$ is the product-moment correlation coefficient R_{jk} between the j th and k th variables, for $j, k = 1, 2, \dots, m$.
- 10: IR – INTEGER *Input*
On entry: the first dimension of the array R as declared in the (sub)program from which G02BAF is called.
Constraint: $IR \geq M$.
- 11: 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, $M < 2$.

IFAIL = 3

On entry, $IX < N$,
or $ISSP < M$,
or $IR < M$.

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 m .

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 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.

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.

```
*      G02BAF Example Program Text
*      Mark 14 Revised.  NAG Copyright 1989.
*      .. Parameters ..
      INTEGER          M, N, IA, ISSP, ICORR
      PARAMETER        (M=3,N=5,IA=N,ISSP=M,ICORR=M)
      INTEGER          NIN, NOUT
      PARAMETER        (NIN=5,NOUT=6)
*      .. Local Scalars ..
      INTEGER          I, IFAIL, J
*      .. Local Arrays ..
      real             A(IA,M), AMEAN(M), CORR(ICORR,M), SSP(ISSP,M),
+                     STD(M)
*      .. External Subroutines ..
      EXTERNAL         G02BAF
*      .. Executable Statements ..
      WRITE (NOUT,*) 'G02BAF Example Program Results'
*      Skip heading in data file
      READ (NIN,*)
      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 G02BAF(N,M,A,IA,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,99996) (I,AMEAN(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,J),J=1,M),I=1,M)
        WRITE (NOUT,*)
        WRITE (NOUT,*) 'Correlation coefficients'
        WRITE (NOUT,99998) (I,I=1,M)
        WRITE (NOUT,99997) (I,(CORR(I,J),J=1,M),I=1,M)
      END IF
      STOP
*
99999 FORMAT (1X,A,I2)
99998 FORMAT (1X,6I12)
99997 FORMAT (1X,I3,3F12.4)
99996 FORMAT (1X,I5,2F11.4)
      END

```

9.2 Program Data

G02BAF Example Program Data

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

9.3 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