NAG Library Routine Document

G02CEF

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

G02CEF takes selected elements from two vectors (typically vectors of means and standard deviations) to form two smaller vectors, and selected rows and columns from two matrices (typically either matrices of sums of squares and cross-products of deviations from means and Pearson product-moment correlation coefficients, or matrices of sums of squares and cross-products about zero and correlation-like coefficients) to form two smaller matrices, allowing reordering of elements in the process.

2 Specification

```
SUBROUTINE G02CEF (N, XBAR, STD, SSP, LDSSP, R, LDR, M, KORDER, XBAR2,      &
                   STD2, SSP2, LDSSP2, R2, LDR2, IFAIL)
INTEGER            N, LDSSP, LDR, M, KORDER(M), LDSSP2, LDR2, IFAIL
REAL (KIND=nag_wp) XBAR(N), STD(N), SSP(LDSSP,N), R(LDR,N), XBAR2(M),       &
                   STD2(M), SSP2(LDSSP2,M), R2(LDR2,M)
```

3 Description

Input to the routine consists of:

(a) A vector of means:

$$(\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots, \bar{x}_n),$$

where n is the number of input variables.

(b) A vector of standard deviations:

$$(s_1, s_2, s_3, \dots, s_n).$$

(c) A matrix of sums of squares and cross-products of deviations from means:

$$\begin{pmatrix} S_{11} & S_{12} & S_{13} & . & . & . & S_{1n} \\ S_{21} & S_{22} & & & & & S_{2n} \\ S_{31} & & & & & & . \\ . & & & & & & . \\ . & & & & & & . \\ . & & & & & & . \\ S_{n1} & S_{n2} & . & . & . & . & S_{nn} \end{pmatrix}.$$

(d) A matrix of correlation coefficients:

$$\begin{pmatrix} R_{11} & R_{12} & R_{13} & . & . & . & R_{1n} \\ R_{21} & R_{22} & & & & & R_{2n} \\ R_{31} & & & & & & . \\ . & & & & & & . \\ . & & & & & & . \\ . & & & & & & . \\ R_{n1} & R_{n2} & . & . & . & . & R_{nn} \end{pmatrix}.$$

(e) The number of variables, m, in the required subset, and their row/column numbers in the input data, $i_1, i_2, i_3, \dots, i_m$,

$$i \le i_k \le n \quad \text{for } k = 1, 2, \dots, m \quad (n \ge 2, m \ge 1 \text{ and } m \le n).$$

New vectors and matrices are output containing the following information:

(i) A vector of means:

$$(\bar{x}_{i_1}, \bar{x}_{i_2}, \bar{x}_{i_3}, \ldots, \bar{x}_{i_m}).$$

(ii) A vector of standard deviations:

$$(s_{i_1}, s_{i_2}, s_{i_3}, \ldots, s_{i_m}).$$

(iii) A matrix of sums of squares and cross-products of deviations from means:

$$\begin{pmatrix} S_{i_1i_1} & S_{i_1i_2} & S_{i_1i_3} & . & . & . & S_{i_1i_m} \\ S_{i_2i_1} & S_{i_2i_2} & & & & & . \\ S_{i_3i_1} & & & & & & . \\ . & & & & & & . \\ . & & & & & & . \\ . & & & & & & . \\ S_{i_mi_1} & S_{i_mi_2} & . & & . & . & S_{i_mi_m} \end{pmatrix}.$$

(iv) A matrix of correlation coefficients:

$$\begin{pmatrix} R_{i_1i_1} & R_{i_1i_2} & R_{i_1i_3} & . & . & . & R_{i_1i_m} \\ R_{i_2i_1} & R_{i_2i_2} & & & & & . \\ R_{i_3i_1} & & & & & & . \\ . & & & & & & . \\ . & & & & & & . \\ . & & & & & & . \\ R_{i_mi_1} & R_{i_mi_2} & . & & . & . & R_{i_mi_m} \end{pmatrix}.$$

Note: for sums of squares of cross-products of deviations about zero and correlation-like coefficients S_{ij} and R_{ij} should be replaced by $\tilde{S}_{ij}$ and $\tilde{R}_{ij}$ in the description of the input and output above.

4 References

None.

5 Parameters

1: N – INTEGER *Input*

On entry: n, the number of variables in the input data.

Constraint: $\mathrm{N} \geq 2$.

2: XBAR(N) – REAL (KIND=nag_wp) array *Input*

On entry: XBAR(i) must be set to $\bar{x}_i$, the mean of variable i, for $i = 1, 2, \ldots, n$.

3: STD(N) – REAL (KIND=nag_wp) array *Input*

On entry: STD(i) must be set to s_i, the standard deviation of variable i, for $i = 1, 2, \ldots, n$.

4: SSP(LDSSP, N) – REAL (KIND=nag_wp) array *Input*

On entry: SSP(i, j) must be set to the sum of cross-products of deviations from means S_{ij} (or about zero, $\tilde{S}_{ij}$) for variables i and j, for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, n$.

5: LDSSP – INTEGER *Input*

On entry: the first dimension of the array SSP as declared in the (sub)program from which G02CEF is called.

Constraint: $\mathrm{LDSSP} \geq \mathrm{N}$.

6: R(LDR, N) – REAL (KIND=nag_wp) array *Input*

On entry: R(i,j) must be set to the Pearson product-moment correlation coefficient R_{ij} (or the correlation-like coefficient, $\tilde{R}_{ij}$) for variables i and j, for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, n$.

7: LDR – INTEGER *Input*

On entry: the first dimension of the array R as declared in the (sub)program from which G02CEF is called.

Constraint: LDR $\geq$ N.

8: M – INTEGER *Input*

On entry: the number of variables m, required in the reduced vectors and matrices.

Constraint: $1 \leq$ M $\leq$ N.

9: KORDER(M) – INTEGER array *Input*

On entry: KORDER(i) must be set to the number of the original variable which is to be the ith variable in the output vectors and matrices, for $i = 1, 2, \ldots, m$.

Constraint: $1 \leq$ KORDER(i) $\leq$ N, for $i = 1, 2, \ldots, m$.

10: XBAR2(M) – REAL (KIND=nag_wp) array *Output*

On exit: the mean of variable i, XBAR(i), where $i =$ KORDER(k), for $k = 1, 2, \ldots, m$. (The array XBAR2 must differ from XBAR and STD.)

11: STD2(M) – REAL (KIND=nag_wp) array *Output*

On exit: the standard deviation of variable i, STD(i), where $i =$ KORDER(k), for $k = 1, 2, \ldots, m$. (The array STD2 must differ from both XBAR and STD.)

12: SSP2(LDSSP2, M) – REAL (KIND=nag_wp) array *Output*

On exit: SSP2(k,l) contains the value of SSP(i,j), where $i =$ KORDER(k) and $j =$ KORDER(l), for $k = 1, 2, \ldots, m$ and $l = 1, 2, \ldots, m$. (The array SSP2 must differ from both SSP and R.)

That is to say: on exit, SSP2(k,l) contains the sum of cross-products of deviations from means S_{ij} (or about zero, $\tilde{S}_{ij}$).

13: LDSSP2 – INTEGER *Input*

On entry: the first dimension of the array SSP2 as declared in the (sub)program from which G02CEF is called.

Constraint: LDSSP2 $\geq$ M.

14: R2(LDR2, M) – REAL (KIND=nag_wp) array *Output*

On exit: R2(k,l) contains the value of R(i,j), where $i =$ KORDER(k) and $j =$ KORDER(l), for $k = 1, 2, \ldots, m$ and $l = 1, 2, \ldots, m$. (The array R2 must differ from both SSP and R.)

That is to say: on exit, R2(k,l) contains the Pearson product-moment coefficient R_{ij} (or the correlation-like coefficient, $\tilde{R}_{ij}$).

15: LDR2 – INTEGER *Input*

On entry: the first dimension of the array R2 as declared in the (sub)program from which G02CEF is called.

Constraint: LDR2 $\geq$ M.

16: IFAIL – INTEGER *Input/Output*

On entry: IFAIL must be set to 0, -1 or 1. If you are unfamiliar with this parameter you should refer to Section 3.3 in the Essential Introduction 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 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.**

On exit: $\text{IFAIL} = 0$ unless the routine detects an error or a warning has been flagged (see Section 6).

6 Error Indicators and Warnings

If on entry $\text{IFAIL} = 0$ or -1, explanatory error messages are output on the current error message unit (as defined by X04AAF).

Errors or warnings detected by the routine:

$\text{IFAIL} = 1$

On entry, $\text{N} < 2$,
or $\text{M} < 1$.

$\text{IFAIL} = 2$

On entry, $\text{N} < \text{M}$.

$\text{IFAIL} = 3$

On entry, $\text{LDSSP} < \text{N}$,
or $\text{LDR} < \text{N}$,
or $\text{LDSSP} < \text{M}$,
or $\text{LDR2} < \text{M}$.

$\text{IFAIL} = 4$

On entry, $\text{KORDER}(i) < 1$,
or $\text{KORDER}(i) > \text{N}$ for some $i = 1, 2, \ldots, m$.

$\text{IFAIL} = -99$

An unexpected error has been triggered by this routine. Please contact NAG.

See Section 3.8 in the Essential Introduction for further information.

$\text{IFAIL} = -399$

Your licence key may have expired or may not have been installed correctly.

See Section 3.7 in the Essential Introduction for further information.

$\text{IFAIL} = -999$

Dynamic memory allocation failed.

See Section 3.6 in the Essential Introduction for further information.

7 Accuracy

Not applicable.

8 Parallelism and Performance

Not applicable.

9 Further Comments

The time taken by G02CEF depends on n and m.

The routine is intended primarily for use when a subset of variables from a larger set of variables is to be used in a regression, and is described accordingly. There is however no reason why the routine should not also be used to select specific rows and columns from vectors and arrays which contain any other non-statistical information; the matrices need not be symmetric.

The routine may be used either with sums of squares and cross-products of deviations from means and Pearson product-moment correlation coefficients in connection with a regression involving a constant, or with sums of squares and cross-products about zero and correlation-like coefficients in connection with a regression with no constant.

10 Example

This example reads in the means, standard deviations, sums of squares and cross-products, and correlation coefficients for four variables. New vectors and matrices are created containing the means, standard deviations, sums of squares and cross-products, and correlation coefficients for the fourth, first and second variables (in that order). Finally these new vectors and matrices are printed.

10.1 Program Text

```
    Program g02cefe

!     G02CEF Example Program Text

!     Mark 25 Release. NAG Copyright 2014.

!     .. Use Statements ..
      Use nag_library, Only: g02cef, nag_wp, x04caf
!     .. Implicit None Statement ..
      Implicit None
!     .. Parameters ..
      Integer, Parameter               :: nin = 5, nout = 6
!     .. Local Scalars ..
      Integer                          :: i, ifail, ldr, ldr2, ldssp, ldssp2,  &
                                          m, n
      Character (80)                   :: fmt
!     .. Local Arrays ..
      Real (Kind=nag_wp), Allocatable  :: r(:,:), r2(:,:), ssp(:,:),          &
                                          ssp2(:,:), std(:), std2(:), xbar(:), &
                                          xbar2(:)
      Integer, Allocatable             :: korder(:)
!     .. Executable Statements ..
      Write (nout,*) 'G02CEF Example Program Results'
      Write (nout,*)

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

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

      ldr = n
      ldr2 = m
      ldssp = n
      ldssp2 = m
      Allocate (r(ldr,n),r2(ldr2,m),ssp(ldssp,n),ssp2(ldssp2,m),std(n), &
        std2(m),xbar(n),xbar2(m),korder(m))

!     Read in the data
```

```
      Read (nin,*) xbar(1:n)
      Read (nin,*) std(1:n)
      Read (nin,*)(ssp(i,1:n),i=1,n)
      Read (nin,*)(r(i,1:n),i=1,n)

!     Read in the reordering vector
      Read (nin,*) korder(1:m)

!     Format for displaying vectors
      Write (fmt,99999) '(1x,a,', n, '(1x,f10.4))'

!     Display data
      Write (nout,fmt) 'Original vector XBAR :   ', xbar(1:n)
      Write (nout,*)
      Write (nout,fmt) 'Original vector STD  :   ', std(1:n)
      Write (nout,*)
      Flush (nout)
      ifail = 0
      Call x04caf('General',' ',n,n,ssp,ldssp,'Original matrix SSP  :',ifail)
      Write (nout,*)
      Flush (nout)
      ifail = 0
      Call x04caf('General',' ',n,n,r,ldr,'Original matrix R    :',ifail)
      Write (nout,*)

!     Calculate summaries on reduced number of variables
      ifail = 0
      Call g02cef(n,xbar,std,ssp,ldssp,r,ldr,m,korder,xbar2,std2,ssp2,ldssp2, &
        r2,ldr2,ifail)

!     Display results
      Write (nout,fmt) 'New vector XBAR2 :   ', xbar2(1:m)
      Write (nout,*)
      Write (nout,fmt) 'New vector STD2  :   ', std2(1:m)
      Write (nout,*)
      Flush (nout)
      ifail = 0
      Call x04caf('General',' ',m,m,ssp2,ldssp2,'New matrix SSP2  :',ifail)
      Write (nout,*)
      Flush (nout)
      ifail = 0
      Call x04caf('General',' ',m,m,r2,ldr2,'New matrix R2    :',ifail)

99999 Format (A,I0,A)
    End Program g02cefe
```

10.2 Program Data

```
G02CEF Example Program Data
4  3                                          :: N, M
  5.8000    2.8000    1.8000    5.4000   :: XBAR
  5.0695    1.9240    2.5884    4.9800   :: STD
102.8000  -29.2000  -14.2000  -57.6000
-29.2000   14.8000   -6.2000    6.4000
-14.2000   -6.2000   28.6000   42.4000
-57.6000    6.4000   42.4000   99.2000   :: End of SSP
  1.0000   -0.7486   -0.2619   -0.5704
 -0.7486    1.0000   -0.3014    0.1670
 -0.2619   -0.3014    1.0000    0.7960
 -0.5704    0.1670    0.7960    1.0000   :: End of R
4  1  2                                  :: KORDER
```

10.3 Program Results

```
 G02CEF Example Program Results

 Original vector XBAR :         5.8000     2.8000     1.8000     5.4000

 Original vector STD  :         5.0695     1.9240     2.5884     4.9800
```

```
Original matrix SSP  :
             1          2          3          4
1    102.8000   -29.2000   -14.2000   -57.6000
2    -29.2000    14.8000    -6.2000     6.4000
3    -14.2000    -6.2000    28.6000    42.4000
4    -57.6000     6.4000    42.4000    99.2000

Original matrix R    :
          1       2       3       4
1   1.0000 -0.7486 -0.2619 -0.5704
2  -0.7486  1.0000 -0.3014  0.1670
3  -0.2619 -0.3014  1.0000  0.7960
4  -0.5704  0.1670  0.7960  1.0000

New vector XBAR2 :        5.4000     5.8000     2.8000

New vector STD2  :        4.9800     5.0695     1.9240

New matrix SSP2  :
             1          2          3
1     99.2000   -57.6000     6.4000
2    -57.6000   102.8000   -29.2000
3      6.4000   -29.2000    14.8000

New matrix R2    :
          1       2       3
1   1.0000 -0.5704  0.1670
2  -0.5704  1.0000 -0.7486
3   0.1670 -0.7486  1.0000
```