

# NAG Library Routine Document

## G02BYF

**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

G02BYF computes a partial correlation/variance-covariance matrix from a correlation or variance-covariance matrix computed by G02BXF.

### 2 Specification

```
SUBROUTINE G02BYF (M, NY, NX, ISZ, R, LDR, P, LDP, WK, IFAIL)
  INTEGER          M, NY, NX, ISZ(M), LDR, LDP, IFAIL
  REAL (KIND=nag_wp) R(LDR,M), P(LDP,NY), WK(NY*NX+NX*(NX+1)/2)
```

### 3 Description

Partial correlation can be used to explore the association between pairs of random variables in the presence of other variables. For three variables,  $y_1$ ,  $y_2$  and  $x_3$ , the partial correlation coefficient between  $y_1$  and  $y_2$  given  $x_3$  is computed as:

$$\frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}},$$

where  $r_{ij}$  is the product-moment correlation coefficient between variables with subscripts  $i$  and  $j$ . The partial correlation coefficient is a measure of the linear association between  $y_1$  and  $y_2$  having eliminated the effect due to both  $y_1$  and  $y_2$  being linearly associated with  $x_3$ . That is, it is a measure of association between  $y_1$  and  $y_2$  conditional upon fixed values of  $x_3$ . Like the full correlation coefficients the partial correlation coefficient takes a value in the range  $(-1, 1)$  with the value 0 indicating no association.

In general, let a set of variables be partitioned into two groups  $Y$  and  $X$  with  $n_y$  variables in  $Y$  and  $n_x$  variables in  $X$  and let the variance-covariance matrix of all  $n_y + n_x$  variables be partitioned into,

$$\begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}.$$

The variance-covariance of  $Y$  conditional on fixed values of the  $X$  variables is given by:

$$\Sigma_{y|x} = \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}.$$

The partial correlation matrix is then computed by standardizing  $\Sigma_{y|x}$ ,

$$\text{diag}(\Sigma_{y|x})^{-\frac{1}{2}}\Sigma_{y|x}\text{diag}(\Sigma_{y|x})^{-\frac{1}{2}}.$$

To test the hypothesis that a partial correlation is zero under the assumption that the data has an approximately Normal distribution a test similar to the test for the full correlation coefficient can be used. If  $r$  is the computed partial correlation coefficient then the appropriate  $t$  statistic is

$$r\sqrt{\frac{n - n_x - 2}{1 - r^2}},$$

which has approximately a Student's  $t$ -distribution with  $n - n_x - 2$  degrees of freedom, where  $n$  is the number of observations from which the full correlation coefficients were computed.

## 4 References

Krzanowski W J (1990) *Principles of Multivariate Analysis* Oxford University Press

Morrison D F (1967) *Multivariate Statistical Methods* McGraw–Hill

Osborn J F (1979) *Statistical Exercises in Medical Research* Blackwell

Snedecor G W and Cochran W G (1967) *Statistical Methods* Iowa State University Press

## 5 Arguments

- 1: M – INTEGER *Input*  
*On entry:* the number of variables in the variance-covariance/correlation matrix given in R.  
*Constraint:*  $M \geq 3$ .
  
- 2: NY – INTEGER *Input*  
*On entry:* the number of  $Y$  variables,  $n_y$ , for which partial correlation coefficients are to be computed.  
*Constraint:*  $NY \geq 2$ .
  
- 3: NX – INTEGER *Input*  
*On entry:* the number of  $X$  variables,  $n_x$ , which are to be considered as fixed.  
*Constraints:*  

$$NX \geq 1;$$

$$NY + NX \leq M.$$
  
- 4: ISZ(M) – INTEGER array *Input*  
*On entry:* indicates which variables belong to set  $X$  and  $Y$ .  

$$ISZ(i) < 0$$
The  $i$ th variable is a  $Y$  variable, for  $i = 1, 2, \dots, M$ .  

$$ISZ(i) > 0$$
The  $i$ th variable is a  $X$  variable.  

$$ISZ(i) = 0$$
The  $i$ th variable is not included in the computations.  
*Constraints:*  
exactly NY elements of ISZ must be  $< 0$ ;  
exactly NX elements of ISZ must be  $> 0$ .
  
- 5: R(LDR, M) – REAL (KIND=nag\_wp) array *Input*  
*On entry:* the variance-covariance or correlation matrix for the  $M$  variables as given by G02BXF. Only the upper triangle need be given.  
**Note:** the matrix must be a full rank variance-covariance or correlation matrix and so be positive definite. This condition is not directly checked by the routine.
  
- 6: LDR – INTEGER *Input*  
*On entry:* the first dimension of the array R as declared in the (sub)program from which G02BYF is called.  
*Constraint:*  $LDR \geq M$ .

- 7: P(LDP, NY) – REAL (KIND=nag\_wp) array *Output*  
*On exit:* the strict upper triangle of P contains the strict upper triangular part of the  $n_y$  by  $n_y$  partial correlation matrix. The lower triangle contains the lower triangle of the  $n_y$  by  $n_y$  partial variance-covariance matrix if the matrix given in R is a variance-covariance matrix. If the matrix given in R is a partial correlation matrix then the variance-covariance matrix is for standardized variables.
- 8: LDP – INTEGER *Input*  
*On entry:* the first dimension of the array P as declared in the (sub)program from which G02BYF is called.  
*Constraint:* LDP  $\geq$  NY.
- 9: WK(NY  $\times$  NX + NX  $\times$  (NX + 1)/2) – REAL (KIND=nag\_wp) array *Workspace*
- 10: 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, M < 3,  
 or NY < 2,  
 or NX < 1,  
 or NY + NX > M,  
 or LDR < M,  
 or LDP < NY.

IFAIL = 2

On entry, there are not exactly NY elements of ISZ < 0,  
 or there are not exactly NX elements of ISZ > 0.

IFAIL = 3

On entry, the variance-covariance/correlation matrix of the  $X$  variables,  $\Sigma_{xx}$ , is not of full rank. Try removing some of the  $X$  variables by setting the appropriate element of ISZ = 0.

IFAIL = 4

Either a diagonal element of the partial variance-covariance matrix,  $\Sigma_{y|x}$ , is zero and/or a computed partial correlation coefficient is greater than one. Both indicate that the matrix input in R was not positive definite.

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

G02BYF computes the partial variance-covariance matrix,  $\Sigma_{y|x}$ , by computing the Cholesky factorization of  $\Sigma_{xx}$ . If  $\Sigma_{xx}$  is not of full rank the computation will fail. For a statement on the accuracy of the Cholesky factorization see F07GDF (DPPTRF).

## 8 Parallelism and Performance

G02BYF is threaded by NAG for parallel execution in multithreaded implementations of the NAG Library.

G02BYF makes calls to BLAS and/or LAPACK routines, which may be threaded within the vendor library used by this implementation. Consult the documentation for the vendor library for further information.

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

Models that represent the linear associations given by partial correlations can be fitted using the multiple regression routine G02DAF.

## 10 Example

Data, given by Osborn (1979), on the number of deaths, smoke (mg/m<sup>3</sup>) and sulphur dioxide (parts/million) during an intense period of fog is input. The correlations are computed using G02BXF and the partial correlation between deaths and smoke given sulphur dioxide is computed using G02BYF. Both correlation matrices are printed using the routine X04CAF.

### 10.1 Program Text

```

Program g02byfe

!      G02BYF Example Program Text

!      Mark 26 Release. NAG Copyright 2016.

!      .. Use Statements ..
      Use nag_library, Only: g02bxf, g02byf, nag_wp, x04caf
!      .. Implicit None Statement ..
      Implicit None
!      .. Parameters ..
      Integer, Parameter          :: nin = 5, nout = 6
!      .. Local Scalars ..

```

```

Integer                                :: i, ifail, ldp, ldr, ldx, lwt, m, n, &
                                   nx, ny
Character (1)                          :: weight
! .. Local Arrays ..
Real (Kind=nag_wp), Allocatable        :: p(:, :), r(:, :), std(:), v(:, :), &
                                   wk(:, :), wt(:, :), x(:, :), xbar(:)
Integer, Allocatable                   :: isz(:)
! .. Executable Statements ..
Write (nout,*) 'G02BYF Example Program Results'
Write (nout,*)
Flush (nout)

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

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

If (weight=='W' .Or. weight=='w') Then
    lwt = n
Else
    lwt = 0
End If
ldp = m
ldr = m
ldx = n
Allocate (p(ldp,m),v(ldr,m),std(m),wk(m*m),wt(lwt),x(ldx,m),xbar(m), &
        isz(m),r(ldr,m))

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

! Read in number of variables and variable flags for partial correlation
! coefficients.
Read (nin,*) ny, nx
Read (nin,*) isz(1:m)

! Calculate correlation matrix
ifail = 0
Call g02bxf(weight,n,m,x,ldx,wt,xbar,std,v,ldr,r,ifail)

! Calculate partial correlation matrix
ifail = 0
Call g02byf(m,ny,nx,isz,r,ldr,p,ldp,wk,ifail)

! Display results
ifail = 0
Call x04caf('Upper','Non-unit',m,m,r,ldr,'Correlation matrix',ifail)
Write (nout,*)
Flush (nout)
ifail = 0
Call x04caf('Upper','Unit',ny,ny,p,ldp,'Partial Correlation matrix', &
        ifail)

End Program g02byfe

```

## 10.2 Program Data

```

G02BYF Example Program Data
'u' 15 3                :: WEIGHT,N,M (G02BXF)
112 0.30 0.09
140 0.49 0.16
143 0.61 0.22
120 0.49 0.14
196 2.64 0.75
294 3.45 0.86
513 4.46 1.34
518 4.46 1.34
430 1.22 0.47
274 1.22 0.47

```

```
255 0.32 0.22
236 0.29 0.23
256 0.50 0.26
222 0.32 0.16
213 0.32 0.16      :: End of X (G02BXF)
  2  1              :: NY,NX
-1 -1  1           :: ISZ
```

### 10.3 Program Results

G02BYF Example Program Results

Correlation matrix

	1	2	3
1	1.0000	0.7560	0.8309
2		1.0000	0.9876
3			1.0000

Partial Correlation matrix

	1	2
1	1.0000	-0.7381
2		1.0000

---