NAG Toolbox for MATLAB

g02by

1 Purpose

g02by computes a partial correlation/variance-covariance matrix from a correlation or variance-covariance matrix computed by g02bx.

2 Syntax

```
[p, ifail] = g02by(ny, nx, isz, r, 'm', m)
```

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:

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

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 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 $\Sigma_{y|x}$,

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

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 Parameters

5.1 Compulsory Input Parameters

1: **ny – int32 scalar**

the number of Y variables, n_y, for which partial correlation coefficients are to be computed.

Constraint: $\mathbf{ny} \geq 2$.

2: **nx – int32 scalar**

The number of X variables, n_x, which are to be considered as fixed.

Constraints:

$\mathbf{nx} \geq 1$;
$\mathbf{ny} + \mathbf{nx} \leq \mathbf{m}$.

3: **isz(m) – int32 array**

Indicates which variables belong to set X and Y.

$\mathbf{isz}(i) < 0$

The ith variable is a Y variable, for $i = 1, 2, \ldots, \mathbf{m}$.

$\mathbf{isz}(i) > 0$

The ith variable is a X variable.

$\mathbf{isz}(i) = 0$

The ith variable is not included in the computations.

Constraints:

exactly **ny** elements of **isz** must be < 0;
exactly **nx** elements of **isz** must be > 0.

4: **r(ldr,m) – double array**

ldr, the first dimension of the array, must be at least **m**.

The variance-covariance or correlation matrix for the **m** variables as given by g02bx. 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 function.

5.2 Optional Input Parameters

1: **m – int32 scalar**

Default: The dimension of the arrays **isz**, **r**. (An error is raised if these dimensions are not equal.)

the number of variables in the variance-covariance/correlation matrix given in **r**.

Constraint: $\mathbf{m} \geq 3$.

5.3 Input Parameters Omitted from the MATLAB Interface

ldr, ldp, wk

5.4 Output Parameters

1: **p(ldp,ny) – double array**

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 correlation matrix then the variance-covariance matrix is for standardized variables.

2: **ifail – int32 scalar**

0 unless the function detects an error (see Section 6).

6 Error Indicators and Warnings

Errors or warnings detected by the function:

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, Σ_{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.

7 Accuracy

g02by computes the partial variance-covariance matrix, $\Sigma_{y|x}$, by computing the Cholesky factorization of Σ_{xx}. If Σ_{xx} is not of full rank the computation will fail. For a statement on the accuracy of the Cholesky factorization see f07gd.

8 Further Comments

Models that represent the linear associations given by partial correlations can be fitted using the multiple regression function g02da.

9 Example

```
ny = int32(2);
nx = int32(1);
isz = [int32(-1);
     int32(-1);
```

```
     int32(1)];
r = [17053.69523809524, 152.8584285714286, 45.94742857142857;
     152.8584285714286, 2.397368571428571, 0.6475057142857144;
     45.94742857142857, 0.6475057142857144, 0.1793171428571428];
[p, ifail] = g02by(ny, nx, isz, r)
```

```
p =
   1.0e+03 *
    5.2803   -0.0007
   -0.0131    0.0001
ifail =
           0
```