NAG Toolbox

nag_correg_corrmat_partial (g02by)

1 Purpose

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

2 Syntax

```
[p, ifail] = nag_correg_corrmat_partial(ny, nx, isz, r, 'm', m)
[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 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** – INTEGER

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

Constraint: **ny** ≥ 2.

2: **nx** – INTEGER

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

Constraints:

nx ≥ 1;
ny + **nx** $\leq$ **m**.

3: **isz**(**m**) – INTEGER array

Indicates which variables belong to set X and Y.

isz$(i) < 0$
The ith variable is a Y variable, for $i = 1, 2, \ldots,$ **m**.

isz$(i) > 0$
The ith variable is a X variable.

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**$)$ – REAL (KIND=nag_wp) array

ldr, the first dimension of the array, must satisfy the constraint $ldr \geq$ **m**.

The variance-covariance or correlation matrix for the **m** variables as given by nag_correg_corrmat (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** – INTEGER

Default: the dimension of the array **isz** and the first dimension of the array **r** and the second dimension of the array **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: **m** ≥ 3.

5.3 Output Parameters

1: **p**$(ldp,$ **ny**$)$ – REAL (KIND=nag_wp) 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 partial correlation matrix then the variance-covariance matrix is for standardized variables.

2: **ifail** – INTEGER

ifail $= 0$ unless the function detects an error (see Section 5).

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.

ifail $= -99$

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

ifail $= -399$

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

ifail $= -999$

Dynamic memory allocation failed.

7 Accuracy

nag_correg_corrmat_partial (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 nag_lapack_dpptrf (f07gd).

8 Further Comments

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

9 Example

Data, given by Osborn (1979), on the number of deaths, smoke (mg/m^3) and sulphur dioxide (parts/million) during an intense period of fog is input. The correlations are computed using nag_correg_corrmat (g02bx) and the partial correlation between deaths and smoke given sulphur dioxide is computed using nag_correg_corrmat_partial (g02by). Both correlation matrices are printed using the function nag_file_print_matrix_real_gen (x04ca).

9.1 Program Text

```
function g02by_example

fprintf('g02by example results\n\n');

x = [ 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];

% Calculate correlation matrix
[xbar, std, v, r, ifail] = g02bx(x);

% Calculate partial correlation matrix
nx = nag_int(1);
ny = nag_int(2);
isz = [nag_int(-1); -1; 1];

[p, ifail] = g02by(ny, nx, isz, r);

mtitle = 'Correlation matrix:';
matrix = 'Upper';
diag   = 'Non-unit';

[ifail] = x04ca( ...
                 matrix, diag, r, mtitle);

fprintf('\n');
mtitle = 'Partial Correlation matrix:';
diag   = 'Unit';

[ifail] = x04ca( ...
                 matrix, diag, p, mtitle);
```

9.2 Program Results

```
g02by example 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
```