NAG Toolbox for MATLAB

g02ce

1 Purpose

g02ce 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 Syntax

```
[xbar2, std2, ssp2, r2, ifail] = g02ce(xbar, std, ssp, r, korder, 'n', n, 'm', m)
```

3 Description

Input to the function consists of:

(a) A vector of means:

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

where n is the number of input variables.

(b) A vector of standard deviations:

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

(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, $i_1, i_2, i_3, \ldots, i_m$,

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

New vectors and matrices are output containing the following information:

(i) A vector of means:

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

(ii) A vector of standard deviations:

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

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

$$\begin{pmatrix} S_{i_1 i_1} & S_{i_1 i_2} & S_{i_1 i_3} & . & . & . & S_{i_1 i_m} \\ S_{i_2 i_1} & S_{i_2 i_2} & & & & & . \\ S_{i_3 i_1} & & & & & & . \\ . & & & & & & . \\ . & & & & & & . \\ . & & & & & & . \\ S_{i_m i_1} & S_{i_m i_2} & . & . & . & . & S_{i_m i_m} \end{pmatrix}.$$

(iv) A matrix of correlation coefficients:

$$\begin{pmatrix} R_{i_1 i_1} & R_{i_1 i_2} & R_{i_1 i_3} & . & . & . & R_{i_1 i_m} \\ R_{i_2 i_1} & R_{i_2 i_2} & & & & & . \\ R_{i_3 i_1} & & & & & & . \\ . & & & & & & . \\ . & & & & & & . \\ . & & & & & & . \\ R_{i_m i_1} & R_{i_m i_2} & . & . & . & . & R_{i_m i_m} \end{pmatrix}.$$

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

5.1 Compulsory Input Parameters

1: **xbar(n) – double array**

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

2: **std(n) – double array**

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

3: **ssp(ldssp,n) – double array**

ldssp, the first dimension of the array, must be at least **n**.

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 $i,j = 1, 2, \ldots, n$.

4: **r(ldr,n) – double array**

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

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 $i,j = 1, 2, \ldots, n$.

5: **korder(m) – int32 array**

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 $i = 1, 2, \ldots, m$.

5.2 Optional Input Parameters

1: **n – int32 scalar**

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

n, the number of variables in the input data.

Constraint: **n** ≥ 2.

2: **m – int32 scalar**

Default: The dimension of the arrays **korder**, **xbar2**, **std2**, **ssp2**, **r2**. (An error is raised if these dimensions are not equal.)

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

Constraint: $1 \leq$ **m** $\leq$ **n**.

5.3 Input Parameters Omitted from the MATLAB Interface

ldssp, ldr, ldssp2, ldr2

5.4 Output Parameters

1: **xbar2(m) – double array**

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**.)

2: **std2(m) – double array**

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**.)

3: **ssp2(ldssp2,m) – double array**

ssp2(k, l) contains the value of **ssp**(i, j), where $i =$ **korder**(k) and $j =$ **korder**(l), for $k, l = 1, 2, \ldots, m$. (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, $\tilde{S}_{ij}$).

4: **r2(ldr2,m) – double array**

r2(k, l) contains the value of **r**(i, j), where $i =$ **korder**(k) and $j =$ **korder**(l), for $k, l = 1, 2, \ldots, m$. (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, $\tilde{R}_{ij}$).

5: **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, $\mathbf{n} < 2$,
or $\mathbf{m} < 1$.

ifail $= 2$

On entry, $\mathbf{n} < \mathbf{m}$.

ifail $= 3$

On entry, **ldssp** $< \mathbf{n}$,
or **ldr** $< \mathbf{n}$,
or **ldssp** $< \mathbf{m}$,
or **ldr2** $< \mathbf{m}$.

ifail $= 4$

On entry, **korder**$(i) < 1$,
or **korder**$(i) > \mathbf{n}$ for some $i = 1, 2, \ldots, m$.

7 Accuracy

Not applicable.

8 Further Comments

The time taken by g02ce depends on n and m.

The function 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 function 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 function 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.

9 Example

```
xbar = [5.8;
     2.8;
     1.8;
     5.4];
std = [5.0695;
     1.924;
     2.5884;
     4.98];
ssp = [102.8, -29.2, -14.2, -57.6;
     -29.2, 14.8, -6.2, 6.4;
     -14.2, -6.2, 28.6, 42.4;
     -57.6, 6.4, 42.4, 99.2];
r = [1, -0.7486, -0.2619, -0.5704;
     -0.7486, 1, -0.3014, 0.167;
     -0.2619, -0.3014, 1, 0.796;
     -0.5704, 0.167, 0.796, 1];
korder = [int32(4);
```

```
     int32(1);
     int32(2)];
[xbar2, std2, ssp2, r2, ifail] = g02ce(xbar, std, ssp, r, korder)
```

```
xbar2 =
    5.4000
    5.8000
    2.8000
std2 =
    4.9800
    5.0695
    1.9240
ssp2 =
   99.2000  -57.6000    6.4000
  -57.6000  102.8000  -29.2000
    6.4000  -29.2000   14.8000
r2 =
    1.0000   -0.5704    0.1670
   -0.5704    1.0000   -0.7486
    0.1670   -0.7486    1.0000
ifail =
           0
```