NAG Toolbox

nag_correg_coeffs_pearson_subset (g02bg)

1 Purpose

nag_correg_coeffs_pearson_subset (g02bg) computes means and standard deviations, sums of squares and cross-products of deviations from means, and Pearson product-moment correlation coefficients for selected variables.

2 Syntax

```
[xbar, std, ssp, r, ifail] = nag_correg_coeffs_pearson_subset(x, kvar, 'n', n, 'm', m, 'nvars', nvars)
[xbar, std, ssp, r, ifail] = g02bg(x, kvar, 'n', n, 'm', m, 'nvars', nvars)
```

Note: the interface to this routine has changed since earlier releases of the toolbox:

At Mark 22: **n** was made optional.

3 Description

The input data consist of n observations for each of m variables, given as an array

$$\left[x_{ij}\right], \quad i = 1,2,\ldots,n(n \geq 2), j = 1,2,\ldots,m(m \geq 2),$$

where x_{ij} is the ith observation on the jth variable, together with the subset of these variables, $v_1, v_2, \ldots, v_p$, for which information is required.

The quantities calculated are:

(a) Means:

$$\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}, \quad j = v_1, v_2, \ldots, v_p.$$

(b) Standard deviations:

$$s_j = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)^2}, \quad j = v_1, v_2, \ldots, v_p.$$

(c) Sums of squares and cross-products of deviations from zero:

$$S_{jk} = \sum_{i=1}^{n}\left(x_{ij} - \bar{x}_j\right)\left(x_{ik} - \bar{x}_k\right), \quad j,k = v_1, v_2, \ldots, v_p.$$

(d) Pearson product-moment correlation coefficients:

$$R_{jk} = \frac{S_{jk}}{\sqrt{S_{jj}S_{kk}}}, \quad j,k = v_1, v_2, \ldots v_p.$$

If S_{jj} or S_{kk} is zero, R_{jk} is set to zero.

4 References

None.

5 Parameters

5.1 Compulsory Input Parameters

1: **x**$(ldx, \mathbf{m})$ – REAL (KIND=nag_wp) array

ldx, the first dimension of the array, must satisfy the constraint $ldx \geq \mathbf{n}$.

x(i, j) must be set to x_{ij}, the value of the ith observation on the jth variable, for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$.

2: **kvar**(**nvars**) – INTEGER array

kvar(j) must be set to the column number in **x** of the jth variable for which information is required, for $j = 1, 2, \ldots, p$.

Constraint: $1 \leq$ **kvar**$(j) \leq$ **m**, for $j = 1, 2, \ldots, p$.

5.2 Optional Input Parameters

1: **n** – INTEGER

Default: the first dimension of the array **x**.

n, the number of observations or cases.

Constraint: **n** ≥ 2.

2: **m** – INTEGER

Default: the second dimension of the array **x**.

m, the number of variables.

Constraint: **m** ≥ 2.

3: **nvars** – INTEGER

Default: the dimension of the array **kvar**.

p, the number of variables for which information is required.

Constraint: $2 \leq$ **nvars** $\leq$ **m**.

5.3 Output Parameters

1: **xbar**(**nvars**) – REAL (KIND=nag_wp) array

The mean value, $\bar{x}_j$, of the variable specified in **kvar**(j), for $j = 1, 2, \ldots, p$.

2: **std**(**nvars**) – REAL (KIND=nag_wp) array

The standard deviation, s_j, of the variable specified in **kvar**(j), for $j = 1, 2, \ldots, p$.

3: **ssp**$(ldssp, \mathbf{nvars})$ – REAL (KIND=nag_wp) array

ssp(j, k) is the cross-product of deviations, S_{jk}, for the variables specified in **kvar**(j) and **kvar**(k), for $j = 1, 2, \ldots, p$ and $k = 1, 2, \ldots, p$.

4: **r**$(ldr, \mathbf{nvars})$ – REAL (KIND=nag_wp) array

r(j, k) is the product-moment correlation coefficient, R_{jk}, between the variables specified in **kvar**(j) and **kvar**(k), for $j = 1, 2, \ldots, p$ and $k = 1, 2, \ldots, p$.

5: **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, **n** < 2.

ifail $= 2$

On entry, **nvars** < 2,
or **nvars** $>$ **m**.

ifail $= 3$

On entry, $ldx <$ **n**,
or $ldssp <$ **nvars**,
or $ldr <$ **nvars**.

ifail $= 4$

On entry, **kvar**$(j) < 1$,
or **kvar**$(j) >$ **m** for some $j = 1, 2, \ldots,$ **nvars**.

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_coeffs_pearson_subset (g02bg) does not use ***additional precision*** arithmetic for the accumulation of scalar products, so there may be a loss of significant figures for large n.

8 Further Comments

The time taken by nag_correg_coeffs_pearson_subset (g02bg) depends on n and p.

The function uses a two pass algorithm.

9 Example

This example reads in a set of data consisting of five observations on each of four variables. The means, standard deviations, sums of squares and cross-products of deviations from means, and Pearson product-moment correlation coefficients for the fourth, first and second variables are then calculated and printed.

9.1 Program Text

```
function g02bg_example

fprintf('g02bg example results\n\n');

x = [ 3,  3,  1,  2;
      6,  4, -1,  4;
      9,  0,  5,  9;
     12,  2,  0,  0;
```

```
     -1,  5,  4, 12];
[n,m] = size(x);
fprintf('Number of variables (columns) = %d\n', m);
fprintf('Number of cases     (rows)    = %d\n\n', n);
disp('Data matrix is:-');
disp(x);

kvar = [nag_int(4); 1; 2];
nvar = size(kvar,1);

[xbar, std, ssp, r, ifail] = g02bg( ...
                                    x, kvar);

fprintf('Variable   Mean     St. dev.\n');
fprintf('%5d%11.4f%11.4f\n',[double(kvar) xbar(1:nvar) std(1:nvar)]');
fprintf('\nSums of squares and cross-products of deviations\n');
disp(ssp(1:nvar,1:nvar))
fprintf('Correlation coefficients\n');
disp(r(1:nvar,1:nvar));
```

9.2 Program Results

```
     g02bg example results

Number of variables (columns) = 4
Number of cases     (rows)    = 5

Data matrix is:-
     3     3     1     2
     6     4    -1     4
     9     0     5     9
    12     2     0     0
    -1     5     4    12

Variable   Mean     St. dev.
    4     5.4000     4.9800
    1     5.8000     5.0695
    2     2.8000     1.9235

Sums of squares and cross-products of deviations
   99.2000  -57.6000    6.4000
  -57.6000  102.8000  -29.2000
    6.4000  -29.2000   14.8000

Correlation coefficients
    1.0000   -0.5704    0.1670
   -0.5704    1.0000   -0.7486
    0.1670   -0.7486    1.0000
```