

## NAG Toolbox

### nag\_correg\_coeffs\_zero (g02bd)

#### 1 Purpose

nag\_correg\_coeffs\_zero (g02bd) computes means and standard deviations of variables, sums of squares and cross-products about zero, and correlation-like coefficients for a set of data.

#### 2 Syntax

```
[xbar, std, sspz, rz, ifail] = nag_correg_coeffs_zero(x, 'n', n, 'm', m)
[xbar, std, sspz, rz, ifail] = g02bd(x, 'n', n, 'm', m)
```

**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 consists of  $n$  observations for each of  $m$  variables, given as an array

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

where  $x_{ij}$  is the  $i$ th observation on the  $j$ th variable.

The quantities calculated are:

(a) Means:

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}, \quad j = 1, 2, \dots, m.$$

(b) Standard deviations:

$$s_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}, \quad j = 1, 2, \dots, m.$$

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

$$\tilde{S}_{jk} = \sum_{i=1}^n x_{ij} x_{ik}, \quad j, k = 1, 2, \dots, m.$$

(d) Correlation-like coefficients:

$$\tilde{R}_{jk} = \frac{\tilde{S}_{jk}}{\sqrt{\tilde{S}_{jj} \tilde{S}_{kk}}}, \quad j, k = 1, 2, \dots, m.$$

If  $\tilde{S}_{jj}$  or  $\tilde{S}_{kk}$  is zero,  $\tilde{R}_{jk}$  is set to zero.

#### 4 References

None.

## 5 Parameters

### 5.1 Compulsory Input Parameters

1: **x**(*ldx*, **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 the value of  $x_{ij}$ , the *i*th observation on the *j*th variable, for  $i = 1, 2, \dots, n$  and  $j = 1, 2, \dots, m$ .

### 5.2 Optional Input Parameters

1: **n** – INTEGER

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

*n*, the number of observations or cases.

*Constraint*:  $\mathbf{n} \geq 2$ .

2: **m** – INTEGER

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

*m*, the number of variables.

*Constraint*:  $\mathbf{m} \geq 2$ .

### 5.3 Output Parameters

1: **xbar**(**m**) – REAL (KIND=nag\_wp) array

**xbar**(*j*) contains the mean value,  $\bar{x}_j$ , of the *j*th variable, for  $j = 1, 2, \dots, m$ .

2: **std**(**m**) – REAL (KIND=nag\_wp) array

The standard deviation,  $s_j$ , of the *j*th variable, for  $j = 1, 2, \dots, m$ .

3: **sspz**(*ldsspz*, **m**) – REAL (KIND=nag\_wp) array

**sspz**(*j*, *k*) is the cross-product about zero,  $\tilde{S}_{jk}$ , for  $j = 1, 2, \dots, m$  and  $k = 1, 2, \dots, m$ .

4: **rz**(*ldrz*, **m**) – REAL (KIND=nag\_wp) array

**rz**(*j*, *k*) is the correlation-like coefficient,  $\tilde{R}_{jk}$ , between the *j*th and *k*th variables, for  $j = 1, 2, \dots, m$  and  $k = 1, 2, \dots, m$ .

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,  $\mathbf{n} < 2$ .

**ifail** = 2

On entry,  $\mathbf{m} < 2$ .

**ifail** = 3

On entry,  $ldx < \mathbf{n}$ ,  
or  $ldsspz < \mathbf{m}$ ,  
or  $ldrz < \mathbf{m}$ .

**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\_zero (g02bd) 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\_zero (g02bd) depends on  $n$  and  $m$ .

The function uses a two-pass algorithm.

## 9 Example

This example reads in a set of data consisting of five observations on each of three variables. The means, standard deviations, sums of squares and cross-products about zero, and correlation-like coefficients for all three variables are then calculated and printed.

### 9.1 Program Text

```
function g02bd_example

fprintf('g02bd example results\n\n');

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

[xbar, std, sspz, rz, ifail] = ...
    g02bd(x);

fprintf('Variable Mean St. dev.\n');
fprintf('%5d%11.4f%11.4f\n', [[1:m]' xbar std]');
fprintf('\nSums of squares and cross-products about zero\n');
disp(sspz)
fprintf('Correlation-like coefficients\n');
disp(rz);
```

## 9.2 Program Results

g02bd example results

Number of variables (columns) = 3  
Number of cases (rows) = 5

Data matrix is:-

2	3	3
4	6	4
9	9	0
0	12	2
12	-1	5

Variable	Mean	St. dev.
1	5.4000	4.9800
2	5.8000	5.0695
3	2.8000	1.9235

Sums of squares and cross-products about zero

245	99	82
99	271	52
82	52	54

Correlation-like coefficients

1.0000	0.3842	0.7129
0.3842	1.0000	0.4299
0.7129	0.4299	1.0000

---