

NAG Toolbox

nag_correg_coeffs_pearson (g02ba)

1 Purpose

nag_correg_coeffs_pearson (g02ba) computes means and standard deviations of variables, sums of squares and cross-products of deviations from means, and Pearson product-moment correlation coefficients for a set of data.

2 Syntax

```
[xbar, std, ssp, r, ifail] = nag_correg_coeffs_pearson(x, 'n', n, 'm', m)
[xbar, std, ssp, r, ifail] = g02ba(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 consist of n observations for each of m variables, given as an array

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

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 of deviations from means:

$$S_{jk} = \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k), \quad j, k = 1, 2, \dots, m.$$

(d) Pearson product-moment correlation coefficients:

$$R_{jk} = \frac{S_{jk}}{\sqrt{S_{jj}S_{kk}}}, \quad j, k = 1, 2, \dots, m.$$

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*, **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 *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

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: **ssp**(*ldssp*, **m**) – REAL (KIND=nag_wp) array

ssp(*j*, *k*) is the cross-product of deviations S_{jk} , for $j = 1, 2, \dots, m$ and $k = 1, 2, \dots, m$.

4: **r**(*ldr*, **m**) – REAL (KIND=nag_wp) array

r(*j*, *k*) is the product-moment correlation coefficient 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 $ldssp < \mathbf{m}$,
 or $ldr < \mathbf{m}$.

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_pearson (g02ba) 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 (g02ba) 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 of deviations from means, and Pearson product-moment correlation coefficients for all three variables are then calculated and printed.

9.1 Program Text

```
function g02ba_example

fprintf('g02ba 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, ssp, r, ifail] = g02ba( ...
                              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 of deviations\n');
disp(ssp)
fprintf('Correlation coefficients\n');
disp(r);
```

9.2 Program Results

g02ba 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 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
