

NAG Toolbox for MATLAB

g02ba

1 Purpose

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] = g02ba(n, x, 'm', m)
```

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: **n** – int32 scalar

n , the number of observations or cases.

Constraint: $n \geq 2$.

2: **x(ldx,m) – double array**

ldx, the first dimension of the array, must be at least **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: **m – int32 scalar**

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

m , the number of variables.

Constraint: $m \geq 2$.

5.3 Input Parameters Omitted from the MATLAB Interface

ldx, **ldssp**, **ldr**

5.4 Output Parameters

1: **xbar(m) – double array**

The mean value, $\bar{x}_j$, of the j th variable, for $j = 1, 2, \dots, m$.

2: **std(m) – double array**

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

3: **ssp(ldssp,m) – double array**

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

4: **r(ldr,m) – double array**

r(j,k) is the product-moment correlation coefficient R_{jk} between the j th and k th variables, for $j, k = 1, 2, \dots, m$.

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, **n** < 2.

ifail = 2

On entry, **m** < 2.

ifail = 3

On entry, **ldx** < **n**,
or **ldssp** < **m**,
or **ldr** < **m**.

7 Accuracy

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 g02ba depends on n and m .

The function uses a two-pass algorithm.

9 Example

```
n = int32(5);
x = [2, 3, 3;
     4, 6, 4;
     9, 9, 0;
     0, 12, 2;
     12, -1, 5];
[xbar, std, ssp, r, ifail] = g02ba(n, x)
```

```
xbar =
    5.4000
    5.8000
    2.8000
std =
    4.9800
    5.0695
    1.9235
ssp =
    99.2000   -57.6000    6.4000
   -57.6000   102.8000   -29.2000
    6.4000   -29.2000   14.8000
r =
    1.0000   -0.5704    0.1670
   -0.5704    1.0000   -0.7486
    0.1670   -0.7486    1.0000
ifail =
    0
```