

NAG Toolbox for MATLAB

g02bd

1 Purpose

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] = g02bd(n, x, 'm', m)
```

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: **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 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 Parameters1: **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, **ldsspz**, **ldrz**

5.4 Output Parameters1: **xbar(m) – double array**

xbar(j) contains 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: **sspz(ldsspz,m) – double array**

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

4: **rz(ldrz,m) – double array**

rz(j,k) is the correlation-like coefficient, $\tilde{R}_{jk}$, 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 **ldsspz** < **m**,
or **ldrz** < **m**.

7 Accuracy

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 g02bd 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, sspz, rz, ifail] = g02bd(n, x)
```

```
xbar =
    5.4000
    5.8000
    2.8000
std =
    4.9800
    5.0695
    1.9235
sspz =
    245    99    82
    99   271    52
    82    52    54
rz =
    1.0000    0.3842    0.7129
    0.3842    1.0000    0.4299
    0.7129    0.4299    1.0000
ifail =
        0
```