NAG Toolbox for MATLAB

g02bg

1 Purpose

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] = g02bg(n, x, kvar, 'm', m, 'nvars', nvars)
```

3 Description

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

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

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:

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

(b) Standard deviations:

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

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

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

(d) Pearson product-moment correlation coefficients:

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

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: **n – int32 scalar**

n, the number of observations or cases.

Constraint: **n** ≥ 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 value of the ith observation on the jth variable, for $i = 1,2,\ldots,n$ and $j = 1,2,\ldots,m$.

3: **kvar(nvars) – int32 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: **m – int32 scalar**

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

m, the number of variables.

Constraint: **m** ≥ 2.

2: **nvars – int32 scalar**

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

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

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

5.3 Input Parameters Omitted from the MATLAB Interface

ldx, ldssp, ldr

5.4 Output Parameters

1: **xbar(nvars) – double array**

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

2: **std(nvars) – double array**

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

3: **ssp(ldssp,nvars) – double array**

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

4: **r(ldr,nvars) – double array**

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

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, $\mathbf{n} < 2$.

ifail $= 2$

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

ifail $= 3$

On entry, $\mathbf{ldx} < \mathbf{n}$,
or $\mathbf{ldssp} < \mathbf{nvars}$,
or $\mathbf{ldr} < \mathbf{nvars}$.

ifail $= 4$

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

7 Accuracy

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 g02bg depends on n and p.

The function uses a two pass algorithm.

9 Example

```
n = int32(5);
x = [3, 3, 1, 2;
     6, 4, -1, 4;
     9, 0, 5, 9;
     12, 2, 0, 0;
     -1, 5, 4, 12];
kvar = [int32(4);
     int32(1);
     int32(2)];
[xbar, std, ssp, r, ifail] = g02bg(n, x, kvar)
```

```
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
```