

NAG Toolbox for MATLAB

g02bk

1 Purpose

g02bk computes means and standard deviations, sums of squares and cross-products about zero, and correlation-like coefficients for selected variables.

2 Syntax

```
[xbar, std, sspz, rz, ifail] = g02bk(n, x, kvar, 'm', m, 'nvars', nvars)
```

3 Description

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

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

where x_{ij} is the i th observation on the j th variable, together with the subset of these variables, $v_1, v_2, \dots, v_p$, for which information is required.

The quantities calculated are:

(a) Means:

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

(b) Standard deviations:

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

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

$$\tilde{S}_{jk} = \sum_{i=1}^n x_{ij}x_{ik}, \quad j, k = v_1, v_2, \dots, v_p.$$

(d) Correlation-like coefficients:

$$\tilde{R}_{jk} = \frac{\tilde{S}_{jk}}{\sqrt{\tilde{S}_{jj}\tilde{S}_{kk}}}, \quad j, k = v_1, v_2, \dots, v_p.$$

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

3: **kvar(nvars) – int32 array**

kvar(j) must be set to the column number in **x** of the j th variable for which information is required, for $j = 1, 2, \dots, p$.

Constraint: $1 \leq \mathbf{kvar}(j) \leq \mathbf{m}$, for $j = 1, 2, \dots, p$.

5.2 Optional Input Parameters1: **m – int32 scalar**

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

m, the number of variables.

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

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: $2 \leq \mathbf{nvars} \leq \mathbf{m}$.

5.3 Input Parameters Omitted from the MATLAB Interface

ldx, ldsspz, ldrz

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

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

2: **std(nvars) – double array**

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

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

ssp(j,k) is the cross-product about zero, $\tilde{S}_{jk}$, for the variables specified in **kvar(j)** and **kvar(k)**, for $j, k = 1, 2, \dots, p$.

4: **rz(ldrz,nvars) – double array**

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

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, **n** < 2.

ifail = 2

On entry, **nvars** < 2,
or **nvars** > **m**.

ifail = 3

On entry, **ldx** < **n**,
or **ldsspz** < **nvars**,
or **ldrz** < **nvars**.

ifail = 4

On entry, **kvar**(*j*) < 1,
or **kvar**(*j*) > **m** for some *j* = 1, 2, ..., **nvars**.

7 Accuracy

g02bk 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 g02bk 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, sspz, rz, ifail] = g02bk(n, x, kvar)
```

```
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	
