

NAG Toolbox for MATLAB

g03za

1 Purpose

g03za produces standardized values (z-scores) for a data matrix.

2 Syntax

```
[z, ifail] = g03za(n, x, nvar, isx, s, e, 'm', m)
```

3 Description

For a data matrix, X , consisting of n observations on p variables, with elements x_{ij} , g03za computes a matrix, Z , with elements z_{ij} such that:

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j}, \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, p,$$

where μ_j is a location shift and σ_j is a scaling factor. Typically, μ_j will be the mean and σ_j will be the standard deviation of the j th variable and therefore the elements in column j of Z will have zero mean and unit variance.

4 References

None.

5 Parameters

5.1 Compulsory Input Parameters

1: **n** – **int32 scalar**

n , the number of observations in the data matrix.

Constraint: $n \geq 1$.

2: **x(ldx,m)** – **double array**

ldx, the first dimension of the array, must be at least **n**.

x(i,j) must contain the i th sample point for the j th variable, x_{ij} , for $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, m$.

3: **nvar** – **int32 scalar**

p , the number of variables to be standardized.

Constraint: $nvar \geq 1$.

4: **isx(m)** – **int32 array**

isx(j) indicates whether or not the observations on the j th variable are included in the matrix of standardized values.

If **isx(j)** $\neq 0$, the observations from the j th variable are included.

If **isx(j)** = 0, the observations from the j th variable are not included.

Constraint: **isx(j)** $\neq 0$ for **nvar** values of j .

5: **s(m) – double array**

If **isx**(*j*) $\neq 0$, **s**(*j*) must contain the scaling (standard deviation), σ_j , for the *j*th variable.

If **isx**(*j*) = 0, **s**(*j*) is not referenced.

Constraint: if **isx**(*j*) $\neq 0$, **s**(*j*) > 0.0, for $j = 1, 2, \dots, \mathbf{m}$.

6: **e(m) – double array**

If **isx**(*j*) $\neq 0$, **e**(*j*) must contain the location shift (mean), μ_j , for the *j*th variable.

If **isx**(*j*) = 0, **e**(*j*) is not referenced.

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

Default: The dimension of the arrays **isx**, **x**. (An error is raised if these dimensions are not equal.)
p, the number of variables in the data array **x**.

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

5.3 Input Parameters Omitted from the MATLAB Interface

ldx, ldz

5.4 Output Parameters1: **z(ldz,nvar) – double array**

The matrix of standardized values (z-scores), *Z*.

2: **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, $\mathbf{n} < 1$,
or $\mathbf{nvar} < 1$,
or $\mathbf{m} < \mathbf{nvar}$,
or $\mathbf{ldx} < \mathbf{n}$,
or $\mathbf{ldz} < \mathbf{n}$.

ifail = 2

On entry, there are not precisely **nvar** elements of **isx** $\neq 0$.

ifail = 3

On entry, **isx**(*j*) $\neq 0$ and **s**(*j*) ≤ 0.0 for some *j*.

7 Accuracy

Standard accuracy is achieved.

8 Further Comments

Means and standard deviations may be obtained using g01aa or g02bx.

9 Example

```
n = int32(4);
x = [15, 0, 1500;
     12, 1, 1000;
     18, 2, 1200;
     14, 3, 500];
nvar = int32(2);
isx = [int32(1);
       int32(0);
       int32(1)];
s = [2.5;
     0;
     420.3];
e = [14.75;
     0;
     1050];
[z, ifail] = g03za(n, x, nvar, isx, s, e)
```

```
z =
    0.1000    1.0707
   -1.1000   -0.1190
    1.3000    0.3569
   -0.3000   -1.3086
ifail =
      0
```