NAG Toolbox

nag_mv_z_scores (g03za)

1 Purpose

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

2 Syntax

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

Note: the interface to this routine has changed since earlier releases of the toolbox:

At Mark 22: **n** was made optional.

3 Description

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

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

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 jth 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: **x**$(ldx, \mathbf{m})$ – REAL (KIND=nag_wp) array

ldx, the first dimension of the array, must satisfy the constraint $ldx \geq \mathbf{n}$.

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

2: **nvar** – INTEGER

p, the number of variables to be standardized.

Constraint: **nvar** ≥ 1.

3: **isx**$(\mathbf{m})$ – INTEGER array

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

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

If $\mathbf{isx}(j) = 0$, the observations from the jth variable are not included.

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

4: **s**(**m**) – REAL (KIND=nag_wp) array

If $\mathbf{isx}(j) \neq 0$, $\mathbf{s}(j)$ must contain the scaling (standard deviation), σ_j, for the jth variable.

If $\mathbf{isx}(j) = 0$, $\mathbf{s}(j)$ is not referenced.

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

5: **e**(**m**) – REAL (KIND=nag_wp) array

If $\mathbf{isx}(j) \neq 0$, $\mathbf{e}(j)$ must contain the location shift (mean), μ_j, for the jth variable.

If $\mathbf{isx}(j) = 0$, $\mathbf{e}(j)$ is not referenced.

5.2 Optional Input Parameters

1: **n** – INTEGER

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

n, the number of observations in the data matrix.

Constraint: $\mathbf{n} \geq 1$.

2: **m** – INTEGER

Default: the dimension of the arrays **isx**, **s**, **e** and the second dimension of the array **x**. (An error is raised if these dimensions are not equal.)

The number of variables in the data array **x**.

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

5.3 Output Parameters

1: **z**(ldz, **nvar**) – REAL (KIND=nag_wp) array

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

2: **ifail** – INTEGER

ifail $= 0$ unless the function detects an error (see Section 5).

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 $ldx < \mathbf{n}$,
or $ldz < \mathbf{n}$.

ifail $= 2$

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

ifail $= 3$

On entry, $\mathbf{isx}(j) \neq 0$ and $\mathbf{s}(j) \leq 0.0$ for some j.

ifail $= -99$

An unexpected error has been triggered by this routine. Please contact NAG.

ifail $= -399$

Your licence key may have expired or may not have been installed correctly.

ifail $= -999$

Dynamic memory allocation failed.

7 Accuracy

Standard accuracy is achieved.

8 Further Comments

Means and standard deviations may be obtained using nag_stat_summary_onevar (g01at) or nag_correg_corrmat (g02bx).

9 Example

A 4 by 3 data matrix is input along with location and scaling values. The first and third columns are scaled and the results printed.

9.1 Program Text

```
     function g03za_example

fprintf('g03za example results\n\n');

x = [15, 0, 1500;
     12, 1, 1000;
     18, 2, 1200;
     14, 3, 500];
nvar = nag_int(2);
isx = [nag_int(1);0;1];

% shift and scaling
s = [ 2.50;     0;    420.3];
e = [14.75;    0;    1050.0];

% Standardize
[z, ifail] = g03za( ...
     x, nvar, isx, s, e);

mtitle = 'Standardized values';
matrix = 'General';
diag   = ' ';
[ifail] = x04ca( ...
                 matrix, diag, z, mtitle);
```

9.2 Program Results

```
     g03za example results

 Standardized values
              1          2
 1      0.1000     1.0707
 2     -1.1000    -0.1190
 3      1.3000     0.3569
 4     -0.3000    -1.3086
```