

NAG Toolbox

nag_smooth_data_order (g10za)

1 Purpose

nag_smooth_data_order (g10za) orders and weights data which is entered unsequentially, weighted or unweighted.

2 Syntax

```
[nord, xord, yord, wtord, rss, ifail] = nag_smooth_data_order(x, y, 'n', n, 'wt', wt)
[nord, xord, yord, wtord, rss, ifail] = g10za(x, y, 'n', n, 'wt', wt)
```

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

At Mark 24: *weight* was removed from the interface; **wt** was made optional.

3 Description

Given a set of observations (x_i, y_i) , for $i = 1, 2, \dots, n$, with corresponding weights w_i , nag_smooth_data_order (g10za) rearranges the observations so that the x_i are in ascending order.

For any equal x_i in the ordered set, say $x_j = x_{j+1} = \dots = x_{j+k}$, a single observation x_j is returned with a corresponding y' and w' , calculated as

$$w' = \sum_{l=0}^k w_{i+l}$$

and

$$y' = \frac{\sum_{l=0}^k w_{i+l} y_{i+l}}{w'}.$$

Observations with zero weight are ignored. If no weights are supplied by you, then unit weights are assumed; that is $w_i = 1$, for $i = 1, 2, \dots, n$.

In addition, the within group sum of squares is computed for the tied observations using West's algorithm (see West (1979)).

4 References

Draper N R and Smith H (1985) *Applied Regression Analysis* (2nd Edition) Wiley

West D H D (1979) Updating mean and variance estimates: An improved method *Comm. ACM* **22** 532–555

5 Parameters

5.1 Compulsory Input Parameters

- 1: **x(n)** – REAL (KIND=nag_wp) array
The values, x_i , for $i = 1, 2, \dots, n$.

2: **y(n)** – REAL (KIND=nag_wp) array

The values y_i , for $i = 1, 2, \dots, n$.

5.2 Optional Input Parameters

1: **n** – INTEGER

Default: the dimension of the arrays **x**, **y**. (An error is raised if these dimensions are not equal.)
n, the number of observations.

Constraint: $n \geq 1$.

2: **wt(:)** – REAL (KIND=nag_wp) array

The dimension of the array **wt** must be at least **n** if *weight* = 'W'

If *weight* = 'W', **wt** must contain the *n* weights. Otherwise **wt** is not referenced and unit weights are assumed.

Constraints:

if *weight* = 'W', $\mathbf{wt}(i) > 0.0$, for $i = 1, 2, \dots, n$;
if *weight* = 'W', $\sum_{i=1}^n \mathbf{wt}(i) > 0$.

5.3 Output Parameters

1: **nord** – INTEGER

The number of distinct observations.

2: **xord(n)** – REAL (KIND=nag_wp) array

The first **nord** elements contain the ordered and distinct x_i .

3: **yord(n)** – REAL (KIND=nag_wp) array

The first **nord** elements contain the values y' corresponding to the values in **xord**.

4: **wtord(n)** – REAL (KIND=nag_wp) array

The first **nord** elements contain the values w' corresponding to the values of **xord** and **yord**.

5: **rss** – REAL (KIND=nag_wp)

The within group sum of squares for tied observations.

6: **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, *weight* \neq 'W' or 'U',
or $n < 1$.

ifail = 2

On entry, *weight* = 'W' and at least one element of **wt** is < 0.0 , or all elements of **wt** are 0.0.

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

For a discussion on the accuracy of the algorithm for computing mean and variance see West (1979).

8 Further Comments

nag_smooth_data_order (g10za) may be used to compute the pure error sum of squares in simple linear regression along with nag_correg_linregm_fit (g02da); see Draper and Smith (1985).

9 Example

A set of unweighted observations are input and nag_smooth_data_order (g10za) used to produce a set of strictly increasing weighted observations.

9.1 Program Text

```
function g10za_example

fprintf('g10za example results\n\n');

x = [1; 3; 5; 5; 3; 4; 9; 6; 9; 9];
y = [4; 4; 1; 2; 5; 3; 4; 9; 7; 4];

% Reorder data
[nord, xord, yord, wtord, rss, ifail] = ...
    g10za(x, y);

% Display results
fprintf('Number of distinct observations = %7d\n', nord);
fprintf('Residual sum of squares          = %13.5f\n\n', rss);
fprintf('%16s%18s%19s\n', 'x', 'y', 'wt');
results = [xord(1:nord) yord(1:nord) wtord(1:nord)];
fprintf('%17.1f%18.1f%18.1f\n', results');
```

9.2 Program Results

```
g10za example results

Number of distinct observations =          6
Residual sum of squares        =          7.00000

           x           y           wt
         1.0         4.0         1.0
         3.0         4.5         2.0
         4.0         3.0         1.0
         5.0         1.5         2.0
         6.0         9.0         1.0
         9.0         5.0         3.0
```
