

NAG Toolbox

nag_correg_ssqmat_update (g02bt)

1 Purpose

nag_correg_ssqmat_update (g02bt) updates the sample means and sums of squares and cross-products, or sums of squares and cross-products of deviations about the mean, for a new observation. The data may be weighted.

2 Syntax

```
[sw, xbar, c, ifail] = nag_correg_ssqmat_update(wt, x, sw, xbar, c, 'mean', mean, 'm', m, 'incx', incx)
[sw, xbar, c, ifail] = g02bt(wt, x, sw, xbar, c, 'mean', mean, 'm', m, 'incx', incx)
```

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

At Mark 24: **mean** was made optional

At Mark 23: **incx** was made optional (default 1).

3 Description

nag_correg_ssqmat_update (g02bt) is an adaptation of West's WV2 algorithm; see West (1979). This function updates the weighted means of variables and weighted sums of squares and cross-products or weighted sums of squares and cross-products of deviations about the mean for observations on m variables X_j , for $j = 1, 2, \dots, m$. For the first $i - 1$ observations let the mean of the j th variable be $\bar{x}_j(i - 1)$, the cross-product about the mean for the j th and k th variables be $c_{jk}(i - 1)$ and the sum of weights be W_{i-1} . These are updated by the i th observation, x_{ij} , for $j = 1, 2, \dots, m$, with weight w_i as follows:

$$W_i = W_{i-1} + w_i, \quad \bar{x}_j(i) = \bar{x}_j(i - 1) + \frac{w_i}{W_i}(x_j - \bar{x}_j(i - 1)), \quad j = 1, 2, \dots, m$$

and

$$c_{jk}(i) = c_{jk}(i - 1) + \frac{w_i}{W_i}(x_j - \bar{x}_j(i - 1))(x_k - \bar{x}_k(i - 1))W_{i-1}, \quad j = 1, 2, \dots, m; k = j, j + 1, 2, \dots, m.$$

The algorithm is initialized by taking $\bar{x}_j(1) = x_{1j}$, the first observation and $c_{ij}(1) = 0.0$.

For the unweighted case $w_i = 1$ and $W_i = i$ for all i .

4 References

Chan T F, Golub G H and Leveque R J (1982) *Updating Formulae and a Pairwise Algorithm for Computing Sample Variances* Compstat, Physica-Verlag

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: **wt** – REAL (KIND=nag_wp)

The weight to use for the current observation, w_i .

For unweighted means and cross-products set **wt** = 1.0. The use of a suitable negative value of **wt**, e.g., $-w_i$ will have the effect of deleting the observation.

- 2: **x**($\mathbf{m} \times \mathbf{incx}$) – REAL (KIND=nag_wp) array

x($(j-1) \times \mathbf{incx} + 1$) must contain the value of the j th variable for the current observation, $j = 1, 2, \dots, m$.

- 3: **sw** – REAL (KIND=nag_wp)

The sum of weights for the previous observations, W_{i-1} .

sw = 0.0

The update procedure is initialized.

sw + **wt** = 0.0

All elements of **xbar** and **c** are set to zero.

Constraint: **sw** \geq 0.0 and **sw** + **wt** \geq 0.0.

- 4: **xbar**(\mathbf{m}) – REAL (KIND=nag_wp) array

If **sw** = 0.0, **xbar** is initialized, otherwise **xbar**(j) must contain the weighted mean of the j th variable for the previous $(i-1)$ observations, $\bar{x}_j(i-1)$, for $j = 1, 2, \dots, m$.

- 5: **c**($(\mathbf{m} \times \mathbf{m} + \mathbf{m})/2$) – REAL (KIND=nag_wp) array

If **sw** \neq 0.0, **c** must contain the upper triangular part of the matrix of weighted sums of squares and cross-products or weighted sums of squares and cross-products of deviations about the mean. It is stored packed form by column, i.e., the cross-product between the j th and k th variable, $k \geq j$, is stored in **c**($k \times (k-1)/2 + j$).

5.2 Optional Input Parameters

- 1: **mean_p** – CHARACTER(1)

Default: 'M'

Indicates whether nag_correg_ssqmat_update (g02bt) is to calculate sums of squares and cross-products, or sums of squares and cross-products of deviations about the mean.

mean = 'M'

The sums of squares and cross-products of deviations about the mean are calculated.

mean = 'Z'

The sums of squares and cross-products are calculated.

Constraint: **mean** = 'M' or 'Z'.

- 2: **m** – INTEGER

Default: the dimension of the array **xbar**.

m , the number of variables.

Constraint: **m** \geq 1.

3: **incx** – INTEGER

Default: 1

The increment of **x**. Two situations are common.

If **incx** = 1, the data values are to be found in consecutive locations in **x**, i.e., in a column.

If **incx** = *ldx*, for some positive integer *ldx*, the data values are to be found as a row of an array with first dimension *ldx*.

Constraint: **incx** > 0.

5.3 Output Parameters

1: **sw** – REAL (KIND=nag_wp)

Contains the updated sum of weights, W_i .

2: **xbar(m)** – REAL (KIND=nag_wp) array

xbar(j) contains the weighted mean of the *j*th variable, $\bar{x}_j(i)$, for $j = 1, 2, \dots, m$.

3: **c((m × m + m)/2)** – REAL (KIND=nag_wp) array

The update sums of squares and cross-products stored as on input.

4: **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, **m** < 1,
or **incx** < 1.

ifail = 2

On entry, **sw** < 0.0.

ifail = 3

On entry, (**sw** + **wt**) < 0.0, the current weight causes the sum of weights to be less than 0.0.

ifail = 4

On entry, **mean** ≠ 'M' or 'Z'.

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 detailed discussion of the accuracy of this method see Chan *et al.* (1982) and West (1979).

8 Further Comments

`nag_correg_ssqmat_update` (g02bt) may be used to update the results returned by `nag_correg_ssqmat` (g02bu).

`nag_correg_ssqmat_to_corrmat` (g02bw) may be used to calculate the correlation matrix from the matrix of sums of squares and cross-products of deviations about the mean and the matrix may be scaled using to produce a variance-covariance matrix.

9 Example

A program to calculate the means, the required sums of squares and cross-products matrix, and the variance matrix for a set of 3 observations of 3 variables.

9.1 Program Text

```
function g02bt_example

fprintf('g02bt example results\n\n');

wt = [0.1300  1.3070  0.3700];
x   = [9.1231  0.9310  0.0009;
       3.7011  0.0900  0.0099;
       4.5230  0.8870  0.0999];
[m,n] = size(x);
cn = (m*(m+1))/2;
m = nag_int(m);

sw   = 0;
xbar = zeros(n,1);
c    = zeros(cn,1);

% Update one observatio at a time
for j = 1:n
    [sw, xbar, c, ifail] = g02bt( ...
                               wt(j), x(:,j), sw, xbar, c);
end

disp('Means');
disp(xbar);

mtitle = 'Sums of squares and cross-products: ';
uplo   = 'Upper';
diag   = 'Non-unit';
[ifail] = x04cc( ...
              uplo, diag, m, c, mtitle);

% Convert the sums of squares and cross-products to a variance matrix
v = c/(sw-1);
fprintf('\n');
mtitle = 'Variance matrix: ';
[ifail] = x04cc( ...
              uplo, diag, m, v, mtitle);
```

9.2 Program Results

```
g02bt example results

Means
    1.3299    0.3334    0.9874

Sums of squares and cross-products:
```

	1	2	3
1	8.7569	3.6978	4.0707
2		1.5905	1.6861
3			1.9297

Variance matrix:

	1	2	3
1	10.8512	4.5822	5.0443
2		1.9709	2.0893
3			2.3912
