NAG Toolbox for MATLAB

g02dc

1   Purpose

g02dc adds or deletes an observation from a general regression model fitted by g02da.

2   Syntax

```
[q, rss, ifail] = g02dc(update, mean, weight, isx, q, ip, x, ix, y, wt,
rss, 'm', m)
```

3   Description

g02da fits a general linear regression model to a data set. You may wish to change the model by either adding or deleting an observation from the data set. g02dc takes the results from g02da and makes the required changes to the vector c and the upper triangular matrix R produced by g02da. The regression coefficients, standard errors and the variance-covariance matrix of the regression coefficients can be obtained from g02dd after all required changes to the data set have been made.

g02da performs a QR decomposition on the (weighted) X matrix of independent variables. To add a new observation to a model with p parameters, the upper triangular matrix R and vector c_1 (the first p elements of c) are augmented by the new observation on independent variables in x^{T} and dependent variable y_{new}. Givens rotations are then used to restore the upper triangular form.

$$\begin{pmatrix} R : c_1 \\ x : y_{\mathrm{new}} \end{pmatrix} \rightarrow \begin{pmatrix} R^* : c_1^* \\ 0 : y_{\mathrm{new}}^* \end{pmatrix}.$$

Note: only R and the upper part of c are updated the remainder of the Q matrix is unchanged.

4   References

Golub G H and Van Loan C F 1996 *Matrix Computations* (3rd Edition) Johns Hopkins University Press, Baltimore

Hammarling S 1985 The singular value decomposition in multivariate statistics *SIGNUM Newsl.* **20 (3)** 2–25

5   Parameters

5.1   Compulsory Input Parameters

1:    **update – string**

Indicates if an observation is to be added or deleted.

update = 'A'

The observation is added.

update = 'D'

The observation is deleted.

Constraint: **update** = 'A' or 'D'.

2:    **mean – string**

Indicates if a mean has been used in the model.

mean $=$ 'M' (Mean)

A mean term or intercept will have been included in the model by g02da.

mean $=$ 'Z'

A model with no mean term or intercept will have been fitted by g02da.

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

3: **weight** − **string**

Indicates if a weight is to be used.

weight $=$ 'U' (Unweighted)

The new observation is unweighted.

weight $=$ 'W' (Weighted)

The new observation is to be weighted and the weight must be supplied in **wt**.

Constraint: **weight** $=$ 'U' or 'W'.

4: **isx(m)** − **int32 array**

If **isx**(j) is greater than 0, the value contained in $\mathbf{x}((j-1)\mathbf{ix}+1)$ is to be included as a value of x^{T}, for $j = 1, 2, \ldots, \mathbf{m}$.

Constraint: if **mean** $=$ 'M', exactly **ip** $- 1$ elements of **isx** must be > 0 and if **mean** $=$ 'Z', exactly **ip** elements of **isx** must be > 0.

5: **q(ldq,ip** $+ \mathbf{1})$ − **double array**

ldq, the first dimension of the array, must be at least **ip**.

Must be array **q** as output by g02da, g02de, g02df, g02ee or a previous call to g02dc.

6: **ip** − **int32 scalar**

The number of linear terms in general linear regression model (including mean if there is one).

Constraint: **ip** ≥ 1.

7: **x(∗)** − **double array**

Note: the dimension of the array **x** must be at least $(\mathbf{m} - 1) \times \mathbf{ix} + 1$.

The **ip** values for the dependent variables of the new observation, x^{T}. The positions will depend on the value of **ix**.

8: **ix** − **int32 scalar**

The increment for elements of **x**. Two situations are common:

ix $= 1$

The values of x are to be chosen from consecutive locations in **x**, i.e., $\mathbf{x}(1), \mathbf{x}(2), \ldots, \mathbf{x}(\mathbf{m})$.

ix $=$ **ldx**

The values of x are to be chosen from a row of a two-dimensional array with first dimension **ldx**, i.e., $\mathbf{x}(1), \mathbf{x}(\mathbf{ldx} + 1), \ldots, \mathbf{x}((\mathbf{m} - 1)\mathbf{ldx} + 1)$.

Constraint: **ix** ≥ 1.

9: **y** − **double scalar**

The value of the dependent variable for the new observation, y_{new}.

10: **wt – double scalar**

If **weight** = 'W', **wt** must contain the weight to be used with the new observation.

If **wt** = 0.0, the observation is not included in the model.

If **weight** = 'U', **wt** is not referenced.

Constraint: if **wt** ≥ 0.0, **weight** = 'W'.

11: **rss – double scalar**

The value of the residual sums of squares for the original set of observations.

Constraint: **rss** ≥ 0.0.

5.2 Optional Input Parameters

1: **m – int32 scalar**

m, the total number of independent variables in the data set.

Constraint: **m** ≥ 1.

5.3 Input Parameters Omitted from the MATLAB Interface

ldq, wk

5.4 Output Parameters

1: **q(ldq,ip + 1) – double array**

The first **ip** elements of the first column of **q** will contain c_1^* the upper triangular part of columns 2 to **ip** + 1 will contain R^* the remainder is unchanged.

2: **rss – double scalar**

The updated values of the residual sums of squares.

Note: this will only be valid if the model is of full rank.

3: **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, **ip** < 1,
or   **ldq** < **ip**,
or   **m** < 1,
or   **ix** < 1,
or   **rss** < 0.0,
or   **update** ≠ 'A' or 'D',
or   **mean** ≠ 'M' or 'Z',
or   **weight** ≠ 'U' or 'W',
or   **mean** = 'M' and there are not exactly **ip** − 1 nonzero values of **isx**,
or   **mean** = 'Z' and there are not exactly **ip** nonzero values of **isx**,

ifail = 2

On entry, **weight** = 'W' and **wt** < 0.0.

ifail $= 3$

> The R matrix could not be updated. This may occur if an attempt is made to delete an observation which was not in the original data set or to add an observation to a R matrix with a zero diagonal element.

ifail $= 4$

> The residual sums of squares cannot be updated. This will occur if the input residual sum of squares is less than the calculated decrease in residual sum of squares when the new observation is deleted.

7 Accuracy

Higher accuracy is achieved by updating the R matrix rather than the traditional methods of updating $\mathbf{x}'X$.

8 Further Comments

Care should be taken with the use of g02dc.

(a) It is possible to delete observations which were not included in the original model.

(b) If several additions/deletions have been performed you are advised to recompute the regression using g02da.

(c) Adding or deleting observations can alter the rank of the model. Such changes will only be detected when a call to g02dd has been made. g02dd should also be used to compute the new residual sum of squares when the model is not of full rank.

g02dc may also be used after g02de, g02df and g02ee.

9 Example

```
update = 'D';
mean = 'Z';
weight = 'U';
isx = [int32(1);
       int32(1);
       int32(1);
       int32(1)];
q = [-72.94999999999999, -2, -0.5, -0.5, -0.5;
            45.48298632955565,  0,  1.6583123951777,  0.4522670168666454,
0.4522670168666454;
            56.11154475279979,  0,  -0.6030226891555274,  1.882937743382544,
0.2896827297511605;
        -42.04198673812751, 0, -0, -0.4963538360840667, -1.860521018838127];
ip = int32(4);
x = [1;
     1;
     1;
     1];
ix = int32(1);
y = 37.89;
wt = 0;
rss = 5274.815851111109;
[qOut, rssOut, ifail] = g02dc(update, mean, weight, isx, q, ip, x, ix, y,
wt, rss)

qOut =
   -62.3596   -1.7321    0.0000   -0.0000    0.0000
    52.3330         0    1.4142    0.0000   -0.0000
    72.0591         0   -0.6030    1.7321    0.0000
   -65.6043         0         0   -0.4964   -1.7321
rssOut =
```

```
   21.7046
ifail =
            0
```