

# NAG Toolbox for MATLAB

## g02df

### 1 Purpose

g02df deletes an independent variable from a general linear regression model.

### 2 Syntax

```
[q, rss, ifail] = g02df(ip, q, indx, rss)
```

### 3 Description

When selecting a linear regression model it is sometimes useful to drop independent variables from the model and to examine the resulting sub-model. g02df updates the  $QR$  decomposition used in the computation of the linear regression model. The  $QR$  decomposition may come from g02da, g02de or a previous call to g02df.

For the general linear regression model with  $p$  independent variables fitted g02da or g02de compute a  $QR$  decomposition of the (weighted) independent variables and form an upper triangular matrix  $R$  and a vector  $c$ . To remove an independent variable  $R$  and  $c$  have to be updated. The column of  $R$  corresponding to the variable to be dropped is removed and the matrix is then restored to upper triangular form by applying a series of Givens rotations. The rotations are then applied to  $c$ . Note only the first  $p$  elements of  $c$  are affected.

The method used means that while the updated values of  $R$  and  $c$  are computed an updated value of  $Q$  from the  $QR$  decomposition is not available so a call to g02de cannot be made after a call to g02df.

g02dd can be used to calculate the parameter estimates,  $\hat{\beta}$ , from the information provided by g02df.

### 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: **ip** – **int32 scalar**

$p$ , the number of independent variables already in the model.

*Constraint:*  $\text{ip} \geq 1$ .

2: **q(ldq,ip + 1)** – **double array**

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

The results of the  $QR$  decomposition as returned by functions g02da, g02dc, g02de, g02ee or previous calls to g02df.

3: **indx** – **int32 scalar**

Indicates which independent variable is to be deleted from the model.

*Constraint:*  $1 \leq \text{indx} \leq \text{ip}$ .

4: **rss – double scalar**

The residual sum of squares for the full regression.

*Constraint:*  $\text{rss} \geq 0.0$ .

**5.2 Optional Input Parameters**

None.

**5.3 Input Parameters Omitted from the MATLAB Interface**

ldq, wk

**5.4 Output Parameters**1: **q(ldq,ip + 1) – double array**

The updated  $QR$  decomposition.

2: **rss – double scalar**

The residual sum of squares with the (**indx**)th variable removed. Note that the residual sum of squares will only be valid if the regression is of full rank, otherwise the residual sum of squares should be obtained using g02dd.

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 **indx** < 1,  
or **indx** > **ip**,  
or **rss** < 0.0.

**ifail = 2**

On entry, a diagonal element of  $R$  is zero.

**7 Accuracy**

There will inevitably be some loss in accuracy in fitting a model by dropping terms from a more complex model rather than fitting it afresh using g02da.

**8 Further Comments**

None.

**9 Example**

```
ip = int32(5);
q = [-27.61466337533981, -3.464101615137754, -12.99038105676658, -
23.72909606369362, ...
```

```

-1.732050807568877, -9.093266739736608;
    5.678919489958151, 0.2542949013547401, 5.979130371550697,
8.354057679970834, 1.505235618012764, 8.634198753323211;
    -1.9627798449884, 0.2542949013547401, 0.1680751103201103, -
3.676917225029723, 0.1291367593463666, 1.878093517739872;
    -0.06450766139237359, 0.2542949013547401, 0.09396625863990909,
0.3650656216718273, 0.8471065054945881, 2.964167523893722;
    -0.2751502181663455, 0.2542949013547401, 0.01985740695970786,
0.4219147717487357, ...
    0.2859917224052277, 0.4467521874922538];
indx = int32(2);
rss = 0.08406649241210093;
[qOut, rssOut, ifail] = g02df(ip, q, indx, rss)

```

```

qOut =
-27.6147    -3.4641   -23.7291    -1.7321    -9.0933   -12.9904
   5.9884    0.2543    9.1274     1.3257    7.1460    5.4725
   0.2703    0.2543    0.1681     1.1147    5.6308    1.5656
  -0.3461    0.2543    0.0940     0.3651   -2.0715   -1.7873
  -0.3582    0.2543    0.0199     0.4219    0.2860   -0.3948
rssOut =
    0.2124
ifail =
    0

```