NAG Fortran Library Routine Document

G02DCF

Note: before using this routine, please read the Users' Note for your implementation to check the interpretation of **bold italicised** terms and other implementation-dependent details.

1   Purpose

G02DCF adds or deletes an observation from a general regression model fitted by G02DAF.

2   Specification

```
      SUBROUTINE G02DCF(UPDATE, MEAN, WEIGHT, M, ISX, Q, LDQ, IP, X, IX, Y,
     1                  WT, RSS, WK, IFAIL)
      INTEGER           M, ISX(M), LDQ, IP, IX, IFAIL
      real              Q(LDQ,IP+1), X(*), Y, WT, RSS, WK(3*IP)
      CHARACTER*1       UPDATE, MEAN, WEIGHT
```

3   Description

G02DAF fits a general linear regression model to a data set. The user may wish to change the model by either adding or deleting an observation from the data set. G02DCF takes the results from G02DAF and makes the required changes to the vector c and the upper triangular matrix R produced by G02DAF. The regression coefficients, standard errors and the variance-covariance matrix of the regression coefficients can be obtained from G02DDF after all required changes to the data set have been made.

G02DAF 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.

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

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

1:   UPDATE – CHARACTER*1   *Input*

On entry: indicates if an observation is to be added or deleted.

If UPDATE = 'A', then the observation is added.

If UPDATE = 'D', then the observation is deleted.

Constraint: UPDATE = 'A' or 'D'.

2:    MEAN – CHARACTER*1   *Input*

On entry: indicates if a mean has been used in the model.

If MEAN = 'M' (Mean), then a mean term or intercept will have been included in the model by G02DAF.

If MEAN = 'Z', then a model with no mean term or intercept will have been fitted by G02DAF.

Constraint: MEAN = 'M' or 'Z'.

3:    WEIGHT – CHARACTER*1   *Input*

On entry: indicates if a weight is to be used.

If WEIGHT = 'U' (Unweighted), the new observation is unweighted.

If WEIGHT = 'W' (Weighted), the new observation is to be weighted and the weight must be supplied in WT.

Constraint: WEIGHT = 'U' or 'W'.

4:    M – INTEGER   *Input*

On entry: the total number of independent variables in the data set.

Constraint: $M \geq 1$.

5:    ISX(M) – INTEGER array   *Input*

On entry: if $ISX(j)$ is greater than 0, then the value contained in $X((j-1)IX+1)$ is to be included as a value of x^T, for $j = 1, 2, \ldots, M$.

Constraint: if MEAN = 'M', then exactly $IP - 1$ elements of ISX must be > 0 and if MEAN = 'Z', then exactly IP elements of ISX must be > 0.

6:    Q(LDQ,IP+1) – ***real*** array   *Input/Output*

On entry: Q must be array Q as output by G02DAF, G02DEF, G02DFF, G02EEF or a previous call to G02DCF.

On exit: 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.

7:    LDQ – INTEGER   *Input*

On entry: the first dimension of the array Q as declared in the (sub)program from which G02DCF is called.

Constraint: $LDQ \geq IP$.

8:    IP – INTEGER   *Input*

On entry: the number of linear terms in general linear regression model (including mean if there is one).

Constraint: $IP \geq 1$.

9:    X(∗) – ***real*** array   *Input*

On entry: the IP values for the dependent variables of the new observation, x^T. The positions will depend on the value of IX.

10:    IX – INTEGER   *Input*

On entry: the increment for elements of X. Two situations are common:

    if $IX = 1$, then the values of x are to be chosen from consecutive locations in X, i.e., X(1),X(2),...,X(M);

if IX = LDX, then the values of x are to be chosen from a row of a two-dimensional array with first dimension LDX, i.e., $X(1), X(LDX + 1), \ldots, X((M - 1)LDX + 1)$.

Constraint: IX $\geq$ 1.

11:   Y – **real**   *Input*

On entry: the value of the dependent variable for the new observation, y_{new}.

12:   WT – **real**   *Input*

On entry: if WEIGHT = 'W', then WT must contain the weight to be used with the new observation.

If WT = 0.0, then the observation is not included in the model.

If WEIGHT = 'U', then WT is not referenced.

Constraint: if WEIGHT = 'W', WT $\geq$ 0.0.

13:   RSS – **real**   *Input/Output*

On entry: the value of the residual sums of squares for the original set of observations.

Constraint: RSS $\geq$ 0.0.

On exit: the updated values of the residual sums of squares.

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

14:   WK(3∗IP) – **real** array   *Workspace*

15:   IFAIL – INTEGER   *Input/Output*

On entry: IFAIL must be set to 0, -1 or 1. Users who are unfamiliar with this parameter should refer to Chapter P01 for details.

On exit: IFAIL = 0 unless the routine detects an error (see Section 6).

For environments where it might be inappropriate to halt program execution when an error is detected, the value -1 or 1 is recommended. If the output of error messages is undesirable, then the value 1 is recommended. Otherwise, for users not familiar with this parameter the recommended value is 0. **When the value -1 or 1 is used it is essential to test the value of IFAIL on exit.**

6   Error Indicators and Warnings

If on entry IFAIL = 0 or -1, explanatory error messages are output on the current error message unit (as defined by X04AAF).

Errors or warnings detected by the routine:

IFAIL = 1

	On entry,	IP $<$ 1,
	or	LDQ $<$ IP,
	or	M $<$ 1,
	or	IX $<$ 1,
	or	RSS $<$ 0.0,
	or	UPDATE $\neq$ 'A' or 'D',
	or	MEAN $\neq$ 'M' or 'Z',
	or	WEIGHT $\neq$ 'U' or 'W',
	or	MEAN = 'M' and there are not exactly IP $-$ 1 non-zero values of ISX,
	or	MEAN = 'Z' and there are not exactly IP non-zero 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 can not 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 $X'X$.

8   Further Comments

Care should be taken with the use of this routine.

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

(b) If several additions/deletions have been performed the user is advised to recompute the regression using G02DAF.

(c) Adding or deleting observations can alter the rank of the model. Such changes will only be detected when a call to G02DDF has been made. G02DDF should also be used to compute the new residual sum of squares when the model is not of full rank.

G02DCF may also be used after G02DEF, G02DFF and G02EEF.

9   Example

A data set consisting of 12 observations with four independent variables is read in and a general linear regression model fitted by G02DAF and parameter estimates printed. The last observation is then dropped and the parameter estimates recalculated, using G02DDF, and printed. Finally a new observation is added and new parameter estimates computed and printed.

9.1   Program Text

Note: the listing of the example program presented below uses ***bold italicised*** terms to denote precision-dependent details. Please read the Users' Note for your implementation to check the interpretation of these terms. As explained in the Essential Introduction to this manual, the results produced may not be identical for all implementations.

```
*     G02DCF Example Program Text
*     Mark 14 Release.  NAG Copyright 1989.
*     .. Parameters ..
      INTEGER         MMAX, NMAX
      PARAMETER       (MMAX=5,NMAX=12)
      INTEGER         NIN, NOUT
      PARAMETER       (NIN=5,NOUT=6)
*     .. Local Scalars ..
      real            RSS, TOL, WTN, YN
      INTEGER         I, IDF, IFAIL, IP, IRANK, J, M, N
      LOGICAL         SVD
      CHARACTER       MEAN, UPDATE, WEIGHT
*     .. Local Arrays ..
*
      real            B(MMAX), COV(MMAX*(MMAX+1)/2), H(NMAX),
     +                P(MMAX*(MMAX+2)), Q(NMAX,MMAX+1), RES(NMAX),
     +                SE(MMAX), WK(5*(MMAX-1)+MMAX*MMAX), WT(NMAX),
```

```
     +                      X(MMAX), XM(NMAX,MMAX), Y(NMAX)
       INTEGER          ISX(MMAX)
*      .. External Subroutines ..
       EXTERNAL         G02DAF, G02DCF, G02DDF
*      .. Executable Statements ..
       WRITE (NOUT,*) 'G02DCF Example Program Results'
*      Skip heading in data file
       READ (NIN,*)
       READ (NIN,*) N, M, WEIGHT, MEAN
       WRITE (NOUT,*)
       IF (N.LE.NMAX .AND. M.LT.MMAX) THEN
          IF (WEIGHT.EQ.'W' .OR. WEIGHT.EQ.'w') THEN
             DO 20 I = 1, N
                READ (NIN,*) (XM(I,J),J=1,M), Y(I), WT(I)
   20        CONTINUE
          ELSE
             DO 40 I = 1, N
                READ (NIN,*) (XM(I,J),J=1,M), Y(I)
   40        CONTINUE
          END IF
          READ (NIN,*) (ISX(J),J=1,M), IP
*         Set tolerance
          TOL = 0.00001e0
          IFAIL = 0
*
*         Fit initial model using G02DAF
          CALL G02DAF(MEAN,WEIGHT,N,XM,NMAX,M,ISX,IP,Y,WT,RSS,IDF,B,SE,
     +                COV,RES,H,Q,NMAX,SVD,IRANK,P,TOL,WK,IFAIL)
*
          WRITE (NOUT,*) 'Results from G02DAF'
          IF (SVD) THEN
             WRITE (NOUT,*)
             WRITE (NOUT,*) 'Model not of full rank'
          END IF
          WRITE (NOUT,99999) 'Residual sum of squares = ', RSS
          WRITE (NOUT,99998) 'Degrees of freedom = ', IDF
          WRITE (NOUT,*)
          WRITE (NOUT,*) 'Variable   Parameter estimate   Standard error'
          WRITE (NOUT,*)
          DO 60 J = 1, IP
             WRITE (NOUT,99997) J, B(J), SE(J)
   60     CONTINUE
   80     READ (NIN,*) UPDATE
          IF (UPDATE.NE.'S' .AND. UPDATE.NE.'s') THEN
             IF (WEIGHT.EQ.'W' .OR. WEIGHT.EQ.'w') THEN
                READ (NIN,*) (X(J),J=1,M), YN, WTN
             ELSE
                READ (NIN,*) (X(J),J=1,M), YN
             END IF
             IFAIL = 0
*
             CALL G02DCF(UPDATE,MEAN,WEIGHT,M,ISX,Q,NMAX,IP,X,1,YN,WTN,
     +                   RSS,WK,IFAIL)
*
             IF (UPDATE.EQ.'A' .OR. UPDATE.EQ.'a') THEN
                WRITE (NOUT,*)
                WRITE (NOUT,*)
     +             'Results from adding an observation using G02DCF'
                N = N + 1
             ELSE IF (UPDATE.EQ.'D' .OR. UPDATE.EQ.'d') THEN
                WRITE (NOUT,*)
                WRITE (NOUT,*)
     +             'Results from dropping an observation using G02DCF'
                N = N - 1
             END IF
             IFAIL = 0
*
             CALL G02DDF(N,IP,Q,NMAX,RSS,IDF,B,SE,COV,SVD,IRANK,P,TOL,WK,
     +                   IFAIL)
*
             WRITE (NOUT,99999) 'Residual sum of squares = ', RSS
```

```
                    WRITE (NOUT,99998) 'Degrees of freedom = ', IDF
                    WRITE (NOUT,*)
                    WRITE (NOUT,*)
         +             'Variable   Parameter estimate   Standard error'
                    WRITE (NOUT,*)
                    DO 100 J = 1, IP
                       WRITE (NOUT,99997) J, B(J), SE(J)
  100             CONTINUE
                    GO TO 80
              END IF
          END IF
          STOP
*
99999 FORMAT (1X,A,e12.4)
99998 FORMAT (1X,A,I4)
99997 FORMAT (1X,I6,2e20.4)
      END
```

9.2   Program Data

```
G02DCF Example Program Data
 12 4  'U' 'Z'
1.0 0.0 0.0 0.0 33.63
0.0 0.0 0.0 1.0 39.62
0.0 1.0 0.0 0.0 38.18
0.0 0.0 1.0 0.0 41.46
0.0 0.0 0.0 1.0 38.02
0.0 1.0 0.0 0.0 35.83
0.0 0.0 0.0 1.0 35.99
1.0 0.0 0.0 0.0 36.58
0.0 0.0 1.0 0.0 42.92
1.0 0.0 0.0 0.0 37.80
0.0 0.0 1.0 0.0 40.43
1.0 1.0 1.0 1.0 37.89
 1   1   1   1   4
'D'
1.0 1.0 1.0 1.0 37.89
'A'
0.0 1.0 0.0 0.0 37.89
'S'
```

9.3   Program Results

```
 G02DCF Example Program Results

 Results from G02DAF
 Residual sum of squares =   0.5275E+04
 Degrees of freedom =    8

 Variable    Parameter estimate    Standard error

     1           0.2072E+02            0.1380E+02
     2           0.1409E+02            0.1624E+02
     3           0.2632E+02            0.1380E+02
     4           0.2260E+02            0.1380E+02

 Results from dropping an observation using G02DCF
 Residual sum of squares =   0.2170E+02
 Degrees of freedom =    7

 Variable    Parameter estimate    Standard error

     1           0.3600E+02            0.1017E+01
     2           0.3701E+02            0.1245E+01
     3           0.4160E+02            0.1017E+01
     4           0.3788E+02            0.1017E+01

 Results from adding an observation using G02DCF
 Residual sum of squares =   0.2223E+02
 Degrees of freedom =    8
```

Variable	Parameter estimate	Standard error
1	0.3600E+02	0.9623E+00
2	0.3730E+02	0.9623E+00
3	0.4160E+02	0.9623E+00
4	0.3788E+02	0.9623E+00