

NAG Fortran Library Routine Document

G02DFF

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

G02DFF deletes an independent variable from a general linear regression model.

2 Specification

```
SUBROUTINE G02DFF(IP, Q, LDQ, INDX, RSS, WK, IFAIL)
INTEGER          IP, LDQ, INDX, IFAIL
real           Q(LDQ,IP+1), RSS, WK(2*IP)
```

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. G02DFF updates the *QR* decomposition used in the computation of the linear regression model. The *QR* decomposition may come from G02DAF, G02DEF or a previous call to G02DFF.

For the general linear regression model with p independent variables fitted G02DAF or G02DEF 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 G02DEF cannot be made after a call to G02DFF.

G02DDF can be used to calculate the parameter estimates, $\hat{\beta}$, from the information provided by G02DFF.

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: IP – INTEGER *Input*
On entry: the number of independent variables already in the model, p .
Constraint: $IP \geq 1$.
- 2: Q(LDQ,IP+1) – **real** array *Input/Output*
On entry: the results of the *QR* decomposition as returned by routines G02DAF, G02DCF, G02DEF, G02EEF or previous calls to G02DFF.
On exit: the updated *QR* decomposition.

- 3: LDQ – INTEGER *Input*
On entry: the first dimension of the array Q as declared in the (sub)program from which G02DFF is called.
Constraint: $LDQ \geq IP$.
- 4: INDX – INTEGER *Input*
On entry: indicates which independent variable is to be deleted from the model.
Constraint: $1 \leq INDX \leq IP$.
- 5: RSS – *real* *Input/Output*
On entry: the residual sum of squares for the full regression.
Constraint: $RSS \geq 0.0$.
On exit: 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 G02DDF.
- 6: WK(2*IP) – *real* array *Workspace*
- 7: 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 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 G02DAF.

8 Further Comments

None.

9 Example

A data set consisting of 12 observations on four independent variables and one dependent variable is read in. The full model, including a mean term, is fitted using G02DAF. The value of INDX is read in and that variable dropped from the regression. The parameter estimates are calculated by G02DDF and printed. This process is repeated until INDX is 0.

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.

```
*      G02DFF 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
INTEGER          I, IDF, IFAIL, INDX, IP, IRANK, J, M, N
LOGICAL          SVD
CHARACTER        MEAN, WEIGHT
*      .. Local Arrays ..
real          B(MMAX), COV((MMAX*MMAX+MMAX)/2), H(NMAX),
+              P(MMAX*(MMAX+2)), Q(NMAX,MMAX+1), RES(NMAX),
+              SE(MMAX), WK(5*(MMAX-1)+MMAX*MMAX), WT(NMAX),
+              X(NMAX,MMAX), Y(NMAX)
INTEGER          ISX(MMAX)
*      .. External Subroutines ..
EXTERNAL         GO2DAF, G02DDF, G02DFF
*      .. Executable Statements ..
WRITE (NOUT,*) 'G02DFF Example Program Results'
*      Skip heading in data file
READ (NIN,*)
READ (NIN,*) N, M, WEIGHT, MEAN
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,*) (X(I,J),J=1,M), Y(I), WT(I)
20    CONTINUE
  ELSE
    DO 40 I = 1, N
      READ (NIN,*) (X(I,J),J=1,M), Y(I)
40    CONTINUE
  END IF
  DO 60 I = 1, M
    ISX(I) = 1
60  CONTINUE
  IP = M
  IF (MEAN.EQ.'M' .OR. MEAN.EQ.'m') IP = IP + 1
*      Set tolerance
TOL = 0.00001e0
IFAIL = 0
*
CALL G02DAF(MEAN,WEIGHT,N,X,NMAX,M,ISX,IP,Y,WT,RSS,IDF,B,SE,
+          COV,RES,H,Q,NMAX,SVD,IRANK,P,TOL,WK,IFAIL)
*
WRITE (NOUT,*)
WRITE (NOUT,*) 'Results from full model'
IF (SVD) THEN
  WRITE (NOUT,*) 'Model not of full rank'
  WRITE (NOUT,*)
END IF
WRITE (NOUT,99999) 'Residual sum of squares = ', RSS
WRITE (NOUT,99998) 'Degrees of freedom = ', IDF
WRITE (NOUT,*)
80  READ (NIN,*) INDX
```

```

      IF (INDX.NE.0) THEN
        IFAIL = 0
*
        CALL G02DFE(IP,Q,NMAX,INDX,RSS,WK,IFAIL)
*
        IP = IP - 1
        IF (IP.EQ.0) THEN
          WRITE (NOUT,*) 'No terms left in model'
        ELSE
          WRITE (NOUT,99998) 'Variable', INDX, ' dropped'
          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,*) 'Parameter estimate    Standard error'
          WRITE (NOUT,*)
          DO 100 J = 1, IP
            WRITE (NOUT,99997) B(J), SE(J)
100      CONTINUE
          GO TO 80
        END IF
      END IF
    END IF
  STOP
*
99999 FORMAT (1X,A,e13.4)
99998 FORMAT (1X,A,I4,A)
99997 FORMAT (1X,e15.4,e20.4)
END

```

9.2 Program Data

G02DFE Example Program Data

```

12 4 'U' 'M'
1.0 1.4 0.0 0.0 4.32
1.5 2.2 0.0 0.0 5.21
2.0 4.5 0.0 0.0 6.49
2.5 6.1 0.0 0.0 7.10
3.0 7.1 0.0 0.0 7.94
3.5 7.7 0.0 0.0 8.53
4.0 8.3 1.0 4.0 8.84
4.5 8.6 1.0 4.5 9.02
5.0 8.8 1.0 5.0 9.27
5.5 9.0 1.0 5.5 9.43
6.0 9.3 1.0 6.0 9.68
6.5 9.2 1.0 6.5 9.83
2
4
0

```

9.3 Program Results

G02DFF Example Program Results

Results from full model

Residual sum of squares = 0.8407E-01
Degrees of freedom = 7

Variable 2 dropped

Residual sum of squares = 0.2124E+00
Degrees of freedom = 8

Parameter estimate Standard error

0.3637E+01	0.1508E+00
0.6126E+00	0.2801E-01
-0.6015E+00	0.4234E+00
0.1671E+00	0.7866E-01

Variable 4 dropped

Residual sum of squares = 0.3322E+00
Degrees of freedom = 9

Parameter estimate Standard error

0.3597E+01	0.1765E+00
0.6209E+00	0.3271E-01
0.2425E+00	0.1724E+00
