NAG 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 (KIND=nag_wp) 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 or 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: p, the number of independent variables already in the model.

Constraint: IP ≥ 1.

2: Q(LDQ,IP $+ 1$) – REAL (KIND=nag_wp) array *Input/Output*

On entry: the results of the QR decomposition as returned by routines G02DAF, G02DCF, G02DEF or 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 (KIND=nag_wp) *Input/Output*

On entry: the residual sum of squares for the full regression.

Constraint: RSS ≥ 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($2 \times$ IP) – REAL (KIND=nag_wp) array *Workspace*

7: IFAIL – INTEGER *Input/Output*

On entry: IFAIL must be set to 0, -1 or 1. If you are unfamiliar with this parameter you should refer to Section 3.3 in the Essential Introduction for details.

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, if you are 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.**

On exit: IFAIL $= 0$ unless the routine detects an error or a warning has been flagged (see Section 6).

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 dataset 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

```
    Program g02dffe

!     G02DFF Example Program Text

!     Mark 24 Release. NAG Copyright 2012.

!     .. Use Statements ..
      Use nag_library, Only: g02daf, g02ddf, g02dff, nag_wp
!     .. Implicit None Statement ..
      Implicit None
!     .. Parameters ..
      Integer, Parameter               :: nin = 5, nout = 6
!     .. Local Scalars ..
      Real (Kind=nag_wp)               :: rss, tol
      Integer                          :: i, idf, ifail, indx, ip, irank, ldq, &
                                          ldx, lwk, lwt, m, n
      Logical                          :: svd
      Character (1)                    :: mean, weight
!     .. Local Arrays ..
      Real (Kind=nag_wp), Allocatable  :: b(:), cov(:), h(:), p(:), q(:,:),    &
                                          res(:), se(:), wk(:), wt(:), x(:,:), &
                                          y(:)
      Integer, Allocatable             :: isx(:)
!     .. Intrinsic Procedures ..
      Intrinsic                        :: max
!     .. Executable Statements ..
      Write (nout,*) 'G02DFF Example Program Results'
      Write (nout,*)

!     Skip heading in data file
      Read (nin,*)
      Read (nin,*) n, m, weight, mean

      If (weight=='W' .Or. weight=='w') Then
        lwt = n
      Else
        lwt = 0
      End If
      ldx = n
      Allocate (x(ldx,m),isx(m),y(n),wt(lwt))

!     Read in data
      If (lwt>0) Then
        Read (nin,*)(x(i,1:m),y(i),wt(i),i=1,n)
      Else
        Read (nin,*)(x(i,1:m),y(i),i=1,n)
      End If

!     Include all variables in the model
      isx(1:m) = 1
      ip = m
      If (mean=='M' .Or. mean=='m') Then
        ip = ip + 1
      End If
      lwk = max(5*(ip-1)+ip*ip,2*ip)
      ldq = n
      Allocate (b(ip),se(ip),cov(ip*(ip+1)/2),res(n),h(n),q(ldq,ip+1),p(2*ip+ &
        ip*ip),wk(lwk))

!     Use suggested value for tolerance
      tol = 0.000001E0_nag_wp

!     Fit general linear regression model
      ifail = 0
      Call g02daf(mean,weight,n,x,ldx,m,isx,ip,y,wt,rss,idf,b,se,cov,res,h,q, &
        ldq,svd,irank,p,tol,wk,ifail)

!     Display results from G02DAF
      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,*)

!     Loop over list of variables to drop
u_lp: Do
        Read (nin,*,Iostat=ifail) indx
        If (ifail/=0) Then
          Exit u_lp
        End If

        If (ip<=0) Then
          Write (nout,*) 'No terms left in model'
          Exit u_lp
        End If

!       Drop variable INDX from the model
        ifail = 0
        Call g02dff(ip,q,ldq,indx,rss,wk,ifail)

        ip = ip - 1
        Write (nout,99998) 'Variable', indx, ' dropped'

!       Calculate parameter estimates etc
        ifail = 0
        Call g02ddf(n,ip,q,ldq,rss,idf,b,se,cov,svd,irank,p,tol,wk,ifail)

!       Display the results for model with variable INDX droped
        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,*)
        Write (nout,99997)(b(i),se(i),i=1,ip)
      End Do u_lp

99999 Format (1X,A,E13.4)
99998 Format (1X,A,I4,A)
99997 Format (1X,E15.4,E20.4)
    End Program g02dffe
```

9.2 Program Data

```
G02DFF Example Program Data
 12 4 'U' 'M'        :: N, M, WEIGHT, MEAN
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 :: End of X, Y
2                    :: Start of variables to drop
4
```

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
```