G02FAF – NAG Fortran Library Routine Document

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

G02FAF calculates two types of standardised residuals and two measures of influence for a linear regression.

2 Specification

```
SUBROUTINE G02FAF(N, IP, NRES, RES, H, RMS, SRES, LDS, IFAIL)
INTEGER        N, IP, NRES, LDS, IFAIL
real           RES(NRES), H(NRES), RMS, SRES(LDS,4)
```

3 Description

For the general linear regression model

$$y = X\beta + \epsilon$$

where y is a vector of length n of the dependent variable,

X is an n by p matrix of the independent variables,

β is a vector of length p of unknown parameters,

and ϵ is a vector of length n of unknown random errors such that $\text{var}\,\epsilon = \sigma^2 I$.

The residuals are given by:

$$r = y - \hat{y} = y - X\hat{\beta}$$

The fitted values, $\hat{y} = X\hat{\beta}$, can be written as Hy for an n by n matrix H. The ith diagonal elements of H, h_i, give a measure of the influence of the ith values of the independent variables on the fitted regression model. The values of r and the h_i are returned by G02DAF.

G02FAF calculates statistics which help to indicate if an observation is extreme and having an undue influence on the fit of the regression model. Two types of standardised residual are calculated:

(a) The ith residual is standardised by its variance when the estimate of σ^2, s^2, is calculated from all the data; known as internal studentization.

$$RI_i = \frac{r_i}{s\sqrt{1 - h_i}}$$

(b) The ith residual is standardised by its variance when the estimate of σ^2, s_{-i}^2 is calculated from the data excluding the ith observation; known as external studentization.

$$RE_i = \frac{r_i}{s_{-i}\sqrt{1 - h_i}} = r_i\sqrt{\frac{n - p - 1}{n - p - RI_i^2}}$$

The two measures of influence are:

(a) Cook's D

$$D_i = \frac{1}{p}RE_i^2\frac{h_i}{1 - h_i}$$

(b) Atkinson's T

$$T_i = |RE_i|\sqrt{\left(\frac{n - p}{p}\right)\left(\frac{h_i}{1 - h_i}\right)}.$$

4 References

[1] Atkinson A C (1981) Two graphical displays for outlying and influential observations in regression *Biometrika* **68** 13–20

[2]   Cook R D and Weisberg S (1982) *Residuals and Influence in Regression* Chapman and Hall

5   Parameters

1:   N — INTEGER   *Input*

On entry: number of observations included in the regression, n.

Constraint: N > IP + 1.

2:   IP — INTEGER   *Input*

On entry: the number of linear parameters estimated in the regression model, p.

Constraint: IP $\geq$ 1.

3:   NRES — INTEGER   *Input*

On entry: the number of residuals.

Constraint: $1 \leq$ NRES $\leq$ N.

4:   RES(NRES) — **real** array   *Input*

On entry: the residuals, r_i.

5:   H(NRES) — **real** array   *Input*

On entry: the diagonal elements of H, h_i, corresponding to the residuals in RES.

Constraint: $0.0 <$ H(i) < 1.0, for $i = 1,2,...,$NRES.

6:   RMS — **real**   *Input*

On entry: the estimate of σ^2 based on all n observations, s^2, i.e., the residual mean square.

Constraint: RMS > 0.0.

7:   SRES(LDS,4) — **real** array   *Output*

On exit: the standardised residuals and influence statistics.

For the observation with residual, r_i, given in RES(i):

> SRES(i, 1) is internally standardised residual, RI_i,
> SRES(i, 2) is externally standardised residual, RE_i,
> SRES(i, 3) is Cook's D statistic, D_i, and
> SRES(i, 4) is Atkinson's T statistic, T_i.

8:   LDS — INTEGER   *Input*

On entry: the dimension of the array SRES as declared in the (sub)program from which G02FAF is called.

Constraint: LDS $\geq$ NRES.

9:   IFAIL — INTEGER   *Input/Output*

On entry: IFAIL must be set to 0, -1 or 1. For users not familiar with this parameter (described in Chapter P01) the recommended value is 0.

On exit: IFAIL = 0 unless the routine detects an error (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 detected by the routine:

IFAIL = 1

> On entry, IP < 1,
>> or N $\leq$ IP + 1,
>> or NRES < 1,
>> or NRES > N,
>> or LDS < NRES,
>> or RMS $\leq$ 0.0.

IFAIL = 2

> On entry, H(i) $\leq$ 0.0 or $\geq$ 1.0, for some $i = 1, 2, \ldots,$ NRES$\ldots$.

IFAIL = 3

> On entry, the value of a residual is too large for the given value of RMS.

7 Accuracy

Accuracy is sufficient for all practical purposes.

8 Further Comments

None.

9 Example

A set of 24 residuals and h_i values from a 11 parameter model fitted to the cloud seeding data considered in Cook and Weisberg [2] are input and the standardised residuals etc calculated and printed for the first 10 observations.

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.

```
*     G02FAF Example Program Text
*     Mark 14 Release.  NAG Copyright 1989.
*     .. Parameters ..
      INTEGER          NMAX
      PARAMETER        (NMAX=24)
      INTEGER          NIN, NOUT
      PARAMETER        (NIN=5,NOUT=6)
*     .. Local Scalars ..
      real             RMS
      INTEGER          I, IFAIL, IP, J, N, NRES
*     .. Local Arrays ..
      real             H(NMAX), RES(NMAX), SRES(NMAX,4)
*     .. External Subroutines ..
      EXTERNAL         G02FAF
*     .. Executable Statements ..
```

```
      WRITE (NOUT,*) 'G02FAF Example Program Results'
*     Skip heading in data file
      READ (NIN,*)
      READ (NIN,*) N, IP, NRES, RMS
      IF (NRES.LT.NMAX) THEN
         DO 20 I = 1, NRES
            READ (NIN,*) RES(I), H(I)
  20     CONTINUE
         IFAIL = 0
*
         CALL G02FAF(N,IP,NRES,RES,H,RMS,SRES,NMAX,IFAIL)
*
         WRITE (NOUT,*)
         WRITE (NOUT,*) '        Internally   Internally'
         WRITE (NOUT,*)
       + 'Obs.    standardized  standardized   Cook''s D   Atkinson''s T'
         WRITE (NOUT,*) '         residuals    residuals'
         WRITE (NOUT,*)
         DO 40 I = 1, NRES
            WRITE (NOUT,99999) I, (SRES(I,J),J=1,4)
  40     CONTINUE
      END IF
      STOP
*
99999 FORMAT (1X,I2,4F13.3)
      END
```

9.2 Program Data

```
G02FAF Example Program Data
 24 11 10 .5798
      0.2660          0.5519
     -0.1387          0.9746
     -0.2971          0.6256
      0.5926          0.3144
     -0.4013          0.4106
      0.1396          0.6268
     -1.3173          0.5479
      1.1226          0.2325
      0.0321          0.4115
     -0.7111          0.3577
      0.3439          0.3342
     -0.4379          0.1673
      0.0633          0.3874
     -0.0936          0.1705
      0.9968          0.3466
      0.0209          0.3743
     -0.4056          0.7527
      0.1396          0.9069
      0.0327          0.2610
      0.2970          0.6256
     -0.2277          0.2485
      0.5180          0.3072
      0.5301          0.5848
     -1.0650          0.4794
```

9.3   Program Results

```
G02FAF Example Program Results

        Internally   Internally
 Obs.  standardized standardized   Cook's D   Atkinson's T
        residuals    residuals

   1      0.522        0.507        0.030        0.611
   2     -1.143       -1.158        4.557       -7.797
   3     -0.638       -0.622        0.062       -0.875
   4      0.940        0.935        0.037        0.689
   5     -0.686       -0.672        0.030       -0.610
   6      0.300        0.289        0.014        0.408
   7     -2.573       -3.529        0.729       -4.223
   8      1.683        1.828        0.078        1.094
   9      0.055        0.053        0.000        0.048
  10     -1.165       -1.183        0.069       -0.960
```