NAG Library Routine Document

G02FAF

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 standardized residuals and two measures of influence for a linear regression.

2 Specification

```
SUBROUTINE G02FAF (N, IP, NRES, RES, H, RMS, SRES, LDSRES, IFAIL)
INTEGER            N, IP, NRES, LDSRES, IFAIL
REAL (KIND=nag_wp) RES(NRES), H(NRES), RMS, SRES(LDSRES,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

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

and 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 standardized residual are calculated:

(i) The ith residual is standardized by its variance when the estimate of σ^2, s^2, is calculated from all the data; this is known as internal Studentization.

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

(ii) The ith residual is standardized by its variance when the estimate of σ^2, s^2_{-i} is calculated from the data excluding the ith observation; this is 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:

(i) Cook's D

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

(ii) Atkinson's T

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

4 References

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

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

5 Parameters

1: N – INTEGER *Input*

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

Constraint: N $>$ IP $+ 1$.

2: IP – INTEGER *Input*

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

Constraint: IP ≥ 1.

3: NRES – INTEGER *Input*

On entry: the number of residuals.

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

4: RES(NRES) – REAL (KIND=nag_wp) array *Input*

On entry: the residuals, r_i.

5: H(NRES) – REAL (KIND=nag_wp) 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 $i = 1, 2, \ldots,$ NRES.

6: RMS – REAL (KIND=nag_wp) *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(LDSRES, 4) – REAL (KIND=nag_wp) array *Output*

On exit: the standardized residuals and influence statistics.

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

SRES$(i, 1)$
Is the internally standardized residual, RI$_i$.

SRES$(i, 2)$
Is the externally standardized residual, RE$_i$.

SRES$(i, 3)$
Is Cook's D statistic, D_i.

SRES$(i, 4)$
Is Atkinson's T statistic, T_i.

8: LDSRES – INTEGER *Input*

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

Constraint: LDSRES $\geq$ NRES.

9: 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 NRES $\leq$ IP + 1,
or NRES $<$ 1,
or NRES $>$ N,
or LDSRES $<$ NRES,
or RMS $\leq$ 0.0.

IFAIL $= 2$

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

IFAIL $= 3$

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

IFAIL $= -99$

An unexpected error has been triggered by this routine. Please contact NAG.

See Section 3.8 in the Essential Introduction for further information.

IFAIL $= -399$

Your licence key may have expired or may not have been installed correctly.

See Section 3.7 in the Essential Introduction for further information.

IFAIL $= -999$

Dynamic memory allocation failed.

See Section 3.6 in the Essential Introduction for further information.

7 Accuracy

Accuracy is sufficient for all practical purposes.

8 Parallelism and Performance

Not applicable.

9 Further Comments

None.

10 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 (1982) are input and the standardized residuals etc calculated and printed for the first 10 observations.

10.1 Program Text

```
    Program g02fafe

!     G02FAF Example Program Text

!     Mark 25 Release. NAG Copyright 2014.

!     .. Use Statements ..
      Use nag_library, Only: g02faf, nag_wp
!     .. Implicit None Statement ..
      Implicit None
!     .. Parameters ..
      Integer, Parameter               :: nin = 5, nout = 6
!     .. Local Scalars ..
      Real (Kind=nag_wp)               :: rms
      Integer                          :: i, ifail, ip, ldsres, n, nres
!     .. Local Arrays ..
      Real (Kind=nag_wp), Allocatable  :: h(:), res(:), sres(:,:)
!     .. Executable Statements ..
      Write (nout,*) 'G02FAF Example Program Results'
      Write (nout,*)

!     Skip heading in data file
      Read (nin,*)

!     Read in the problem size
      Read (nin,*) n, ip, nres, rms

      ldsres = nres
      Allocate (res(nres),h(nres),sres(ldsres,4))

!     Read in the data
      Read (nin,*)(res(i),h(i),i=1,nres)

!     Calculate standardised residuals
      ifail = 0
      Call g02faf(n,ip,nres,res,h,rms,sres,ldsres,ifail)

!     Display results
      Write (nout,*) '        Internally    Internally'
      Write (nout,*) &
        'Obs.   standardized  standardized   Cook''s D   Atkinson''s T'
      Write (nout,*) '         residuals     residuals'
```

```
      Write (nout,*)
      Write (nout,99999)(i,sres(i,1:4),i=1,nres)

99999 Format (1X,I2,4F13.3)
    End Program g02fafe
```

10.2 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
```

10.3 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
```