NAG Library Routine Document

F04YAF

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

F04YAF returns elements of the estimated variance-covariance matrix of the sample regression coefficients for the solution of a linear least squares problem.

The routine can be used to find the estimated variances of the sample regression coefficients.

2 Specification

```
SUBROUTINE F04YAF (JOB, P, SIGMA, A, LDA, SVD, IRANK, SV, CJ, WORK, IFAIL)

INTEGER            JOB, P, LDA, IRANK, IFAIL
REAL (KIND=nag_wp) SIGMA, A(LDA,P), SV(P), CJ(P), WORK(P)
LOGICAL            SVD
```

3 Description

The estimated variance-covariance matrix C of the sample regression coefficients is given by

$$C = \sigma^2 \left(X^{\mathrm{T}} X\right)^{-1}, \qquad X^{\mathrm{T}} X \qquad \text{nonsingular},$$

where $X^{\mathrm{T}} X$ is the normal matrix for the linear least squares regression problem

$$\min : \|y - Xb\|_2, \tag{1}$$

σ^2 is the estimated variance of the residual vector $r = y - Xb$, and X is an n by p observation matrix.

When $X^{\mathrm{T}} X$ is singular, C is taken to be

$$C = \sigma^2 \left(X^{\mathrm{T}} X\right)^{\dagger},$$

where $\left(X^{\mathrm{T}} X\right)^{\dagger}$ is the pseudo-inverse of $X^{\mathrm{T}} X$; this assumes that the minimal least squares solution of (1) has been found.

The diagonal elements of C are the estimated variances of the sample regression coefficients, b.

The routine can be used to find either the diagonal elements of C, or the elements of the jth column of C, or the upper triangular part of C.

This routine must be preceded by a routine that returns either the upper triangular matrix U of the QU factorization of X or of the Cholesky factorization of $X^{\mathrm{T}} X$, or the singular values and right singular vectors of X. In particular this routine can be preceded by one of the routines F04JGF or F08KAF (DGELSS), which return the parameters IRANK, SIGMA, A and SV in the required form. F04JGF returns the parameter SVD, but when this routine is used following routine F08KAF (DGELSS) the parameter SVD should be set to .TRUE.. The parameter P of this routine corresponds to the parameter N in routines F04JGF and F08KAF (DGELSS).

4 References

Anderson T W (1958) *An Introduction to Multivariate Statistical Analysis* Wiley

Lawson C L and Hanson R J (1974) *Solving Least-squares Problems* Prentice–Hall

5 Parameters

1: JOB – INTEGER *Input*

On entry: specifies which elements of C are required.

JOB $= -1$
The upper triangular part of C is required.

JOB $= 0$
The diagonal elements of C are required.

JOB > 0
The elements of column JOB of C are required.

Constraint: $-1 \leq \text{JOB} \leq \text{P}$.

2: P – INTEGER *Input*

On entry: p, the order of the variance-covariance matrix C.

Constraint: $\text{P} \geq 1$.

3: SIGMA – REAL (KIND=nag_wp) *Input*

On entry: σ, the standard error of the residual vector given by

$$\sigma = \sqrt{r^{\mathrm{T}} r/(n-k)}, \quad n > k$$

$$\sigma = 0, \qquad n = k,$$

where k is the rank of X.

Constraint: $\text{SIGMA} \geq 0.0$.

4: A(LDA,P) – REAL (KIND=nag_wp) array *Input/Output*

On entry: if SVD = .FALSE., A must contain the upper triangular matrix U of the QU factorization of X, or of the Cholesky factorization of $X^{\mathrm{T}}X$; elements of the array below the diagonal need not be set.

If SVD = .TRUE., A must contain the first k rows of the matrix V^{T}, where k is the rank of X and V is the right-hand orthogonal matrix of the singular value decomposition of X. Thus the ith row must contain the ith right-hand singular vector of X.

On exit: if $\text{JOB} \geq 0$, A is unchanged.

If $\text{JOB} = -1$, A contains the upper triangle of the symmetric matrix C.

If SVD = .TRUE., elements of the array below the diagonal are used as workspace.

If SVD = .FALSE., they are unchanged.

5: LDA – INTEGER *Input*

On entry: the first dimension of the array A as declared in the (sub)program from which F04YAF is called.

Constraints:

if SVD = .FALSE. or $\text{JOB} = -1$, $\text{LDA} \geq \text{P}$;
if SVD = .TRUE. and $\text{JOB} \geq 0$, $\text{LDA} \geq \max(1, \text{IRANK})$.

6: SVD – LOGICAL *Input*

On entry: must be .TRUE. if the least squares solution was obtained from a singular value decomposition of X. SVD must be .FALSE. if the least squares solution was obtained from either a

QU factorization of X or a Cholesky factorization of $X^{\mathrm{T}}X$. In the latter case the rank of X is assumed to be p and so is applicable only to full rank problems with $n \geq p$.

7: IRANK – INTEGER *Input*

On entry: if SVD = .TRUE., IRANK must specify the rank k of the matrix X.

If SVD = .FALSE., IRANK is not referenced and the rank of X is assumed to be p.

Constraint: $0 <$ IRANK $\leq \min(n, \mathrm{P})$.

8: SV(P) – REAL (KIND=nag_wp) array *Input*

On entry: if SVD = .TRUE., SV must contain the first IRANK singular values of X.

If SVD = .FALSE., SV is not referenced.

9: CJ(P) – REAL (KIND=nag_wp) array *Output*

On exit: if JOB = 0, CJ returns the diagonal elements of C.

If JOB $= j > 0$, CJ returns the jth column of C.

If JOB $= -1$, CJ is not referenced.

10: WORK(P) – REAL (KIND=nag_wp) array *Workspace*

If JOB > 0, WORK is not referenced.

11: 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, P < 1,
or SIGMA < 0.0,
or JOB < -1,
or JOB $>$ P,
or SVD = .TRUE. and (IRANK < 0 or IRANK $>$ P)
or (JOB ≥ 0 and LDA $< \max(1, \mathrm{IRANK})$)
or (JOB $= -1$ and LDA $<$ P)),
or SVD = .FALSE. and LDA $<$ P.

IFAIL = 2

On entry, SVD = .TRUE. and IRANK = 0.

IFAIL = 3

On entry, SVD = .FALSE. and overflow would occur in computing an element of C. The upper triangular matrix U must be very nearly singular.

IFAIL = 4

On entry, SVD = .TRUE. and one of the first IRANK singular values is zero. Either the first IRANK singular values or IRANK must be incorrect.

overflow

If overflow occurs then either an element of C is very large, or more likely, either the rank, or the upper triangular matrix, or the singular values or vectors have been incorrectly supplied.

7 Accuracy

The computed elements of C will be the exact covariances of a closely neighbouring least squares problem, so long as a numerically stable method has been used in the solution of the least squares problem.

8 Further Comments

When JOB $= -1$ the time taken by F04YAF is approximately proportional to pk^2, where k is the rank of X. When JOB $= 0$ and SVD = .FALSE., the time taken by the routine is approximately proportional to pk^2, otherwise the time taken is approximately proportional to pk.

9 Example

This example finds the estimated variances of the sample regression coefficients (the diagonal elements of C) for the linear least squares problem

$$\min r^{\mathrm{T}}r, \qquad \text{where } r = y - Xb \qquad \text{and}$$

$$X = \begin{pmatrix} 0.6 & 1.2 & 3.9 \\ 5.0 & 4.0 & 2.5 \\ 1.0 & -4.0 & -5.5 \\ -1.0 & -2.0 & -6.5 \\ -4.2 & -8.4 & -4.8 \end{pmatrix}, \qquad b = \begin{pmatrix} 3.0 \\ 4.0 \\ -1.0 \\ -5.0 \\ -1.0 \end{pmatrix},$$

following a solution obtained by F04JGF. See the routine document for F04JGF for further information.

9.1 Program Text

```
    Program f04yafe

!     F04YAF Example Program Text

!     Mark 24 Release. NAG Copyright 2012.

!     .. Use Statements ..
      Use nag_library, Only: f04jgf, f04yaf, nag_wp
!     .. Implicit None Statement ..
      Implicit None
!     .. Parameters ..
      Integer, Parameter                 :: nin = 5, nout = 6
!     .. Local Scalars ..
      Real (Kind=nag_wp)                 :: sigma, tol
      Integer                            :: i, ifail, irank, job, lda, lwork, n, p
      Logical                            :: svd
!     .. Local Arrays ..
      Real (Kind=nag_wp), Allocatable    :: a(:,:), cj(:), work(:), y(:)
```

```
!     .. Executable Statements ..
      Write (nout,*) 'F04YAF Example Program Results'
!     Skip heading in data file
      Read (nin,*)
      Read (nin,*) n, p
      lda = n
      lwork = 4*p
      Allocate (a(lda,p),cj(p),work(lwork),y(n))
      tol = 5.0E-4_nag_wp
      Read (nin,*)(a(i,1:p),i=1,n)
      Read (nin,*) y(1:n)

!     ifail: behaviour on error exit
!             =0 for hard exit, =1 for quiet-soft, =-1 for noisy-soft
      ifail = 0
      Call f04jgf(n,p,a,lda,y,tol,svd,sigma,irank,work,lwork,ifail)

      Write (nout,*)
      Write (nout,99999) 'SIGMA =', sigma, '  Rank =', irank, '  SVD =', svd
      Write (nout,*)
      Write (nout,*) 'Solution vector'
      Write (nout,99998) y(1:p)
      job = 0

      ifail = 0
      Call f04yaf(job,p,sigma,a,lda,svd,irank,work,cj,work(p+1),ifail)

      Write (nout,*)
      Write (nout,*) 'Estimated variances of regression coefficients'
      Write (nout,99998) cj(1:p)

99999 Format (1X,A,F9.4,A,I3,A,L3)
99998 Format (1X,7F9.4)
    End Program f04yafe
```

9.2 Program Data

```
F04YAF Example Program Data
   5   3                          : n, p
 0.60  1.20  3.90
 5.00  4.00  2.50
 1.00 -4.00 -5.50
-1.00 -2.00 -6.50
-4.20 -8.40 -4.80                 : matrix A
 3.0   4.0  -1.0  -5.0  -1.0      : vector Y
```

9.3 Program Results

```
 F04YAF Example Program Results

 SIGMA =   0.4123  Rank =  3  SVD =  F

 Solution vector
    0.9533  -0.8433   0.9067

 Estimated variances of regression coefficients
    0.0106   0.0093   0.0045
```