NAG Fortran Library Routine Document

G04DAF

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

G04DAF computes sum of squares for a user defined contrast between means.

2 Specification

```
 SUBROUTINE G04DAF(NT, TMEAN, IREP, RMS, RDF, NC, CT, LDCT, EST, TABLE,
1                  LDT, TOL, USETX, TX, IFAIL)
 INTEGER           NT, IREP(NT), NC, LDCT, LDT, IFAIL
 real              TMEAN(NT), RMS, RDF, CT(LDCT,NC), EST(NC),
1                  TABLE(LDT,5), TOL, TX(NT)
 LOGICAL           USETX
```

3 Description

In the analysis of designed experiments the first stage is to compute the basic analysis of variance table, the estimate of the error variance (the residual or error mean square), $\hat{\sigma}^2$, and the (variance ratio) F-statistic for the t treatments. If this F test is significant then the second stage of the analysis is to explore which treatments are significantly different.

If there is a structure to the treatments then this may lead to hypotheses that can be defined before the analysis and tested using linear contrasts. For example, if the treatments were three different fixed temperatures, say 18, 20 and 22, and an uncontrolled temperature (denoted by N) then the following contrasts might be of interest.

	18	20	22	N
(a)	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	-1
(b)	-1	0	1	0

The first represents the average difference between the controlled temperatures and the uncontrolled temperature. The second represents the linear effect of an increasing fixed temperature.

For a randomised complete block design or a completely randomised design, let the treatment means be $\hat{\tau}_i$, $i = 1, 2, \ldots, t$, and let the jth contrast be defined by λ_{ij}, $i = 1, 2, \ldots, t$, then the estimate of the contrast is simply:

$$\Lambda_j = \sum_{i=1}^{t} \hat{\tau}_i \lambda_{ij}$$

and the sum of squares for the contrast is:

$$SS_j = \frac{\Lambda_j^2}{\sum_{i=1}^{t} \lambda_{ij}^2 / n_i} \quad (1)$$

where n_i is the number of observations for the ith treatment. Such a contrast has one degree of freedom so that the appropriate F-statistic is $SS_j/\hat{\sigma}^2$.

The two contrasts λ_{ij} and $\lambda_{ij'}$ are orthogonal if $\sum_{i=1}^{t} \lambda_{ij}\lambda_{ij'} = 0$ and the contrast λ_{ij} is orthogonal to the overall mean if $\sum_{i=1}^{t} \lambda_{ij} = 0$. In practice these sums will be tested against a small quantity, ϵ. If each of a set of contrasts is orthogonal to the mean and they are all mutually orthogonal then the contrasts provide a partition of the treatment sum of squares into independent components. Hence the resulting F-tests are independent.

If the treatments come from a design in which treatments are not orthogonal to blocks then the sum of squares for a contrast is given by:

$$SS_j = \frac{\Lambda_j \Lambda_j^*}{\sum_{i=1}^{t} \lambda_{ij}^2 / n_i} \tag{2}$$

where

$$\Lambda_j^* = \sum_{i=1}^{t} \tau_i^* \lambda_{ij}$$

with τ_i^*, for $i = 1, 2, \ldots, t$, being adjusted treatment means computed by first eliminating blocks then computing the treatment means from the block adjusted observations without taking into account the non-orthogonality between treatments and blocks. For further details see John (1987) and Morgan (1993).

4 References

Cochran W G and Cox G M (1957) *Experimental Designs* Wiley

John J A (1987) *Cyclic Designs* Chapman and Hall

Morgan G W (1993) Analysis of variance using the NAG Fortran Library: Examples from Cochran and Cox *NAG Technical Report TR 3/93* NAG Ltd, Oxford

Winer B J (1970) *Statistical Principles in Experimental Design* McGraw-Hill

5 Parameters

1: NT – INTEGER *Input*

On entry: the number of treatment means, t.

Constraint: NT ≥ 2.

2: TMEAN(NT) – ***real*** array *Input*

On entry: the treatment means, $\hat{\tau}_i$, for $i = 1, 2, \ldots, t$.

3: IREP(NT) – INTEGER array *Input*

On entry: the replication for each treatment mean, n_i, for $i = 1, 2, \ldots, t$.

4: RMS – ***real*** *Input*

On entry: the residual mean square, $\hat{\sigma}^2$.

Constraint: RMS > 0.0.

5: RDF – ***real*** *Input*

On entry: the residual degrees of freedom.

Constraint: RDF ≥ 1.0.

6: NC – INTEGER *Input*

On entry: the number of contrasts.

Constraint: NC $\geq$ 1.

7: CT(LDCT,NC) – ***real*** array *Input*

On entry: the columns of CT must contain the NC contrasts, that is CT(i,j) must contain λ_{ij}, for $i = 1, 2, \ldots, t$; $j = 1, 2, \ldots,$ NC.

8: LDCT – INTEGER *Input*

On entry: the first dimension of the array CT as declared in the (sub)program from which G04DAF is called.

Constraint: LDCT $\geq$ NT.

9: EST(NC) – ***real*** array *Output*

On exit: the estimates of the contrast, Λ_j, for $j = 1, 2, \ldots,$ NC.

10: TABLE(LDT,5) – ***real*** array *Output*

On exit: the rows of the analysis of variance table for the contrasts. For each row column 1 contains the degrees of freedom, column 2 contains the sum of squares, column 3 contains the mean square, column 4 the F-statistic and column 5 the significance level for the contrast. Note that the degrees of freedom are always one and so the mean square equals the sum of squares.

11: LDT – INTEGER *Input*

On entry: the first dimension of the array TABLE as declared in the (sub)program from which G04DAF is called.

Constraint: LDT $\geq$ NC.

12: TOL – ***real*** *Input*

On entry: the tolerance, ϵ used to check if the contrasts are orthogonal and if they are orthogonal to the mean. If TOL $\leq$ 0.0 the value ***machine precision*** is used.

13: USETX – LOGICAL *Input*

On entry: if USETX = .TRUE. the means τ_i^* are provided in TX and the formula (2) is used instead of formula (1). If USETX = .FALSE. formula (1) is used and TX is not referenced.

14: TX(NT) – ***real*** array *Input*

On entry: if USETX = .TRUE. TX must contain the means τ_i^*, for $i = 1, 2, \ldots, t$.

15: 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, because for this routine the values of the output parameters may be useful even if IFAIL $\neq$ 0 on exit, the recommended value is -1. **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, NC < 1,
or NT < 2,
or LDCT < NT,
or LDT < NC,
or RMS ≤ 0.0,
or RDF < 1.0.

IFAIL = 2

On entry, a contrast is not orthogonal to the mean,
or at least two contrasts are not orthogonal.

If IFAIL = 2 full results are returned but they should be interpreted with care.

7 Accuracy

The computations are stable.

8 Further Comments

If the treatments have a factorial structure G04CAF should be used and if the treatments have no structure the means can be compared using G04DBF.

9 Example

The data is given in Morgan (1993) and is for a completely randomised experiment on potato scab with seven treatments representing amounts of sulphur applied, whether the application was in spring or autumn and a control treatment. The one-way anova is computed using G02BBF. Two contrasts are analysed, one comparing the control with use of sulphur, the other comparing spring with autumn application.

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.

```
*     G04DAF Example Program Text
*     Mark 17 Release. NAG Copyright 1995.
*     .. Parameters ..
      INTEGER          NIN, NOUT
      PARAMETER        (NIN=5,NOUT=6)
      INTEGER          NMAX, NTMAX, NBMAX
      PARAMETER        (NMAX=32,NTMAX=7,NBMAX=1)
*     .. Local Scalars ..
      real             GMEAN, RDF, RMS, TOL
      INTEGER          I, IFAIL, IRDF, J, LDT, N, NBLOCK, NC, NT
*     .. Local Arrays ..
      real             BMEAN(NBMAX), C(NTMAX,NTMAX), CT(NTMAX,NTMAX),
     +                 EF(NTMAX), EST(NTMAX), R(NMAX), TABLE(NTMAX+4,5),
     +                 TMEAN(NTMAX), TX(NTMAX), WK(NTMAX*NTMAX+NTMAX),
     +                 Y(NMAX)
      INTEGER          IREP(NTMAX), IT(NMAX)
      CHARACTER*11     NAMES(NTMAX)
*     .. External Subroutines ..
      EXTERNAL         G04BBF, G04DAF
*     .. Executable Statements ..
```

```
      WRITE (NOUT,*) 'G04DAF Example Program Results'
*     Skip heading in data file
      READ (NIN,*)
      READ (NIN,*) N, NT
      IF (N.LE.NMAX .AND. NT.LE.NTMAX) THEN
         READ (NIN,*) (Y(I),I=1,N)
         READ (NIN,*) (IT(I),I=1,N)
         TOL = 0.000005e0
         IRDF = 0
         NBLOCK = 1
         LDT = NTMAX + 4
         IFAIL = -1
         CALL G04BBF(N,Y,NBLOCK,NT,IT,GMEAN,BMEAN,TMEAN,TABLE,LDT,C,
     +               NTMAX,IREP,R,EF,TOL,IRDF,WK,IFAIL)
         WRITE (NOUT,*)
         WRITE (NOUT,*) ' ANOVA table'
         WRITE (NOUT,*)
         WRITE (NOUT,*)
     +     '  Source        df         SS           MS           F',
     +     '        Prob'
         WRITE (NOUT,*)
         WRITE (NOUT,99998) ' Treatments', (TABLE(2,J),J=1,5)
         WRITE (NOUT,99998) ' Residual  ', (TABLE(3,J),J=1,3)
         WRITE (NOUT,99998) ' Total     ', (TABLE(4,J),J=1,2)
*
         RMS = TABLE(3,3)
         RDF = TABLE(3,1)
         READ (NIN,*) NC
         WRITE (NOUT,*)
         WRITE (NOUT,*) ' Orthogonal Contrasts'
         WRITE (NOUT,*)
         DO 20 I = 1, NC
            READ (NIN,*) (CT(J,I),J=1,NT)
            READ (NIN,99999) NAMES(I)
   20    CONTINUE
         CALL G04DAF(NT,TMEAN,IREP,RMS,RDF,NC,CT,NTMAX,EST,TABLE(5,1),
     +               LDT,TOL,.FALSE.,TX,IFAIL)
         DO 40 I = 1, NC
            WRITE (NOUT,99998) NAMES(I), (TABLE(I+4,J),J=1,5)
   40    CONTINUE
      END IF
      STOP
*
99999 FORMAT (A)
99998 FORMAT (A,3X,F3.0,2X,2(F10.1,2X),F10.3,2X,F9.4)
      END
```

9.2 Program Data

```
G04DAF Example Program Data

32 7

12 10 24 29 30 18 32 26
9 9 16 4 30 7 21 9 16 10 18 18
18 24 12 19 10 4 4 5 17 7 16 17

1 1 1 1 1 1 1 1
2 2 2 2 3 3 3 3 4 4 4 4
5 5 5 5 6 6 6 6 7 7 7 7

2

6 -1 -1 -1 -1 -1 -1
 Cntl v S
0 1 -1 1 -1 1 -1
 Spring v A
```

9.3 Program Results

```
G04DAF Example Program Results

 ANOVA table

  Source        df          SS          MS          F        Prob

Treatments    6.       972.3       162.1       3.608     0.0103
Residual     25.      1122.9        44.9
Total        31.      2095.2

 Orthogonal Contrasts

Cntl v S      1.       518.0       518.0      11.533     0.0023
Spring v A    1.       228.2       228.2       5.080     0.0332
```