NAG Fortran Library Routine Document

G04CAF

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

G04CAF computes an analysis of variance table and treatment means for a complete factorial design.

2 Specification

```
      SUBROUTINE G04CAF(N, Y, NFAC, LFAC, NBLOCK, INTER, IRDF, MTERM, TABLE,
     1                  ITOTAL, TMEAN, MAXT, E, IMEAN, SEMEAN, BMEAN, R, IWK,
     2                  IFAIL)
      INTEGER           N, NFAC, LFAC(NFAC), NBLOCK, INTER, IRDF, MTERM,
     1                  ITOTAL, MAXT, IMEAN(MTERM), IWK(N+3*NFAC), IFAIL
      real              Y(N), TABLE(MTERM,5), TMEAN(MAXT), E(MAXT),
     1                  SEMEAN(MTERM), BMEAN(NBLOCK+1), R(N)
```

3 Description

An experiment consists of a collection of units, or plots, to which a number of treatments are applied. In a factorial experiment the effects of several different sets of conditions are compared, e.g., three different temperatures, T_1, T_2 and T_3, and two different pressures, P_1 and P_2. The conditions are known as factors and the different values the conditions take are known as levels. In a factorial experiment the experimental treatments are the combinations of all the different levels of all factors. e.g.,

$$T_1P_1, \quad T_2P_1, \quad T_3P_1$$

$$T_1P_2, \quad T_2P_2, \quad T_3P_2$$

The effect of a factor averaged over all other factors is known as a main effect and the effect of a combination of some of the factors averaged over all other factors is known as an interaction. This can be represented by a linear model. In the above example if the response was y_{ijk} for the kth replicate of the ith level of T and the jth level of P the linear model would be:

$$y_{ijk} = \mu + t_i + p_j + \gamma_{ij} + e_{ijk}$$

where μ is the overall mean, t_i is the main effect of T, p_j is the main effect of P, γ_{ij} is the $T \times P$ interaction and e_{ijk} is the random error term. In order to find unique estimates constraints are placed on the parameters estimates. For the example here these are:

$$\sum_{i=1}^{3} \hat{t}_i = 0,$$

$$\sum_{j=1}^{2} \hat{p}_j = 0,$$

$$\sum_{i=1}^{3} \hat{\gamma}_{ij} = 0, \quad \text{for } j = 1, 2 \text{ and}$$

$$\sum_{j=1}^{2} \hat{\gamma}_{ij} = 0, \quad \text{for } i = 1, 2, 3,$$

where ˆ denotes the estimate.

If there is variation in the experimental conditions, e.g., in an experiment on the production of a material different batches of raw material may be used, or the experiment may be carried out on different days, then plots that are similar are grouped together into blocks. For a balanced complete factorial experiment all the treatment combinations occur the same number of times in each block.

G04CAF computes the analysis of variance (ANOVA) table by sequentially computing the totals and means for an effect from the residuals computed when previous effects have been removed. The effect sum of squares is the sum of squared totals divided by the number of observations per total. The means are then subtracted from the residuals to compute a new set of residuals. At the same time the means for the original data are computed. When all effects are removed the residual sum of squares is computed from the residuals. Given the sums of squares an ANOVA table is then computed along with standard errors for the difference in treatment means.

The data for G04CAF has to be in standard order given by the order of the factors. Let there be k factors, $f_1, f_2, \ldots, f_k$ in that order with levels $l_1, l_2, \ldots, l_k$ respectively. Standard order requires the levels of factor f_1 are in order $1, 2, \ldots, l_1$ and within each level of f_1 the levels of f_2 are in order $1, 2, \ldots, l_2$ and so on.

For an experiment with blocks the data is for block 1 then for block 2 etc. Within each block the data must be arranged so that the levels of factor f_1 are in order $1, 2, \ldots, l_1$ and within each level of f_1 the levels of f_2 are in order $1, 2, \ldots, l_2$ and so on. Any within block replication of treatment combinations must occur within the levels of f_k.

The ANOVA table is given in the following order. For a complete factorial experiment the first row is for blocks, if present, then the main effects of the factors in their order, e.g., f_1 followed by f_2 etc. These are then followed by all the two factor interactions then all the three factor interactions etc., the last two rows being for the residual and total sums of squares. The interactions are arranged in lexical order for the given order. For example, for the three factor interactions for a five factor experiment the 10 interactions would be in the following order:

$$
\begin{array}{c}
f_1\,f_2\,f_3\\
f_1\,f_2\,f_4\\
f_1\,f_2\,f_5\\
f_1\,f_3\,f_4\\
f_1\,f_3\,f_5\\
f_1\,f_4\,f_5\\
f_2\,f_3\,f_4\\
f_2\,f_3\,f_5\\
f_2\,f_4\,f_5\\
f_3\,f_4\,f_5
\end{array}
$$

4 References

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

Davis O L (1978) *The Design and Analysis of Industrial Experiments* Longman

John J A and Quenouille M H (1977) *Experiments: Design and Analysis* Griffin

5 Parameters

1: N – INTEGER *Input*

On entry: the number of observations.

Constraints:

$N \geq 4$.
N must be a multiple of NBLOCK if NBLOCK > 1.
N must be a multiple of the number of treatment combinations, that is a multiple of $\prod_{i=1}^{k} \mathrm{LFAC}(i)$.

2: Y(N) – ***real*** array *Input*

On entry: the observations in standard order, see Section 3.

3: NFAC – INTEGER *Input*

On entry: the number of factors, k.

Constraint: NFAC ≥ 1.

4: LFAC(NFAC) – INTEGER array *Input*

On entry: LFAC(i) must contain the number of levels for the ith factor, for $i = 1, 2, \ldots, k$.

Constraint: LFAC(i) ≥ 2, for $i = 1, 2, \ldots, k$.

5: NBLOCK – INTEGER *Input*

On entry: the number of blocks. If there are no blocks, set NBLOCK=0 or 1.

Constraints:

NBLOCK ≥ 0.
If NBLOCK ≥ 2, N/NBLOCK must be a multiple of the number of treatment combinations, that is a multiple of $\prod_{i=1}^{k}$ LFAC(i).

6: INTER – INTEGER *Input*

On entry: the maximum number of factors in an interaction term. If no interaction terms are to be computed, set INTER=0 or 1.

Constraint: $0 \leq$ INTER $\leq$ NFAC.

7: IRDF – INTEGER *Input*

On entry: the adjustment to the residual and total degrees of freedom. The total degrees of freedom are set to N $-$ IRDF and the residual degrees of freedom adjusted accordingly. For examples of the use of IRDF see Section 8.

Constraint: IRDF ≥ 0.

8: MTERM – INTEGER *Input*

On entry: the maximum number of terms in the analysis of variance table, see Section 8.

Constraint: MTERM must be large enough to contain the terms specified by NFAC, INTER and NBLOCK. If the routine exits with IFAIL ≥ 2, the required minimum value of MTERM is returned in ITOTAL.

9: TABLE(MTERM,5) – ***real*** array *Output*

On exit: the first ITOTAL rows of TABLE contain the analysis of variance table. The first column contains the degrees of freedom, the second column contains the sum of squares, the third column (except for the row corresponding to the total sum of squares) contains the mean squares, i.e., the sums of squares divided by the degrees of freedom, and the fourth and fifth columns contain the F ratio and significance level, respectively (except for rows corresponding to the total sum of squares, and the residual sum of squares). All other cells of the table are set to zero.

The first row corresponds to the blocks and is set to zero if there are no blocks. The ITOTALth row corresponds to the total sum of squares for Y and the (ITOTAL $-$ 1)th row corresponds to the residual sum of squares. The central rows of the table correspond to the main effects followed by the interaction if specified by INTER. The main effects are in the order specified by LFAC and the interactions are in lexical order, see Section 3.

10: ITOTAL – INTEGER *Output*

On exit: the row in TABLE corresponding to the total sum of squares. The number of treatment effects is ITOTAL $-$ 3.

11: TMEAN(MAXT) – ***real*** array *Output*

On exit: the treatment means. The position of the means for an effect is given by the index in IMEAN. For a given effect the means are in standard order, see Section 3.

12: MAXT – INTEGER *Input*

On entry: the maximum number of treatment means to be computed, see Section 8. If the value of MAXT is too small for the required analysis then the minimum number is returned in IMEAN(1).

Constraint: MAXT must be large enough for the number of means specified by LFAC and INTER; if INTER=NFAC then $\text{MAXT} \geq \prod_{i=1}^{k}(\text{LFAC}(i)+1)-1$.

13: E(MAXT) – ***real*** array *Output*

On exit: the estimated effects in the same order as for the means in TMEAN.

14: IMEAN(MTERM) – INTEGER array *Output*

On exit: indicates the position of the effect means in TMEAN. The effect means corresponding to the first treatment effect in the ANOVA table are stored in TMEAN(1) up to TMEAN(IMEAN(1)). Other effect means corresponding to the ith treatment effect, $i = 1, 2, \ldots, \text{ITOTAL} - 3$, are stored in TMEAN(IMEAN($i-1$) + 1) up to TMEAN(IMEAN(i)).

15: SEMEAN(MTERM) – ***real*** array *Output*

On exit: the standard error of the difference between means corresponding to the ith treatment effect in the ANOVA table.

16: BMEAN(NBLOCK+1) – ***real*** array *Output*

On exit: BMEAN(1) contains the grand mean, if $\text{NBLOCK} > 1$, BMEAN(2) up to BMEAN($\text{NBLOCK}+1$) contain the block means.

17: R(N) – ***real*** array *Output*

On exit: the residuals.

18: IWK(N+3*NFAC) – INTEGER array *Workspace*

19: 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, for users 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.**

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, N < 4,
or NFAC < 1,
or NBLOCK < 0,
or INTER < 0,
or INTER > NFAC,
or IRDF < 0.

IFAIL = 2

On entry, LFAC(i) $\leq$ 1, for some $i = 1, 2, \ldots,$ NFAC,
or the value of MAXT is too small,
or the value of MTERM is too small,
or NBLOCK > 1 and N is not a multiple of NBLOCK,
or the number of plots per block is not a multiple of the number of treatment combinations.

IFAIL = 3

On entry, the values of Y are constant.

IFAIL = 4

There are no degrees of freedom for the residual or the residual sum of squares is zero. In either case the standard errors and F-statistics cannot be computed.

7 Accuracy

The block and treatment sums of squares are computed from the block and treatment residual totals. The residuals are updated as each effect is computed and the residual sum of squares computed directly from the residuals. This avoids any loss of accuracy in subtracting sums of squares.

8 Further Comments

The number of rows in the ANOVA table and the number of treatment means are given by the following formulae.

Let there be k factors with levels l_i for $i = 1, 2, \ldots, k$. and let t be the maximum number of terms in an interaction then the number of rows in the ANOVA tables is:

$$\sum_{i=1}^{t} \binom{k}{i} + 3.$$

The number of treatment means is:

$$\sum_{i=1}^{t} \prod_{j \in S_i} l_j,$$

where S_i is the set of all combinations of the k factors i at a time.

To estimate missing values the Healy and Westmacott procedure or its derivatives may be used, see John and Quenouille (1977). This is an iterative procedure in which estimates of the missing values are adjusted by subtracting the corresponding values of the residuals. The new estimates are then used in the analysis of variance. This process is repeated until convergence. A suitable initial value may be the grand mean. When using this procedure IRDF should be set to the number of missing values plus one to obtain the correct degrees of freedom for the residual sum of squares.

For analysis of covariance the residuals are obtained from an analysis of variance of both the response variable and the covariates. The residuals from the response variable are then regressed on the residuals from the covariates using, say, G02CBF or G02DAF. The coefficients obtained from the regression can be examined for significance and used to produce an adjusted dependent variable using the original response variable and covariate. An approximate adjusted analysis of variance table can then be produced by using the adjusted dependent variable. In this case IRDF should be set to one plus the number of fitted covariates.

For designs such as Latin squares one more of the blocking factors has to be removed in a preliminary analysis before the final analysis. This preliminary analysis can be performed using G04BBF or a prior call to G04CAF if the data is reordered between calls. The residuals from the preliminary analysis are then input to G04CAF. In these cases IRDF should be set to the difference between N and the residual degrees of freedom from preliminary analysis. Care should be taken when using this approach as there is no check on the orthogonality of the two analyses.

9 Example

The data, given by John and Quenouille (1977), is for the yield of turnips for a factorial experiment with two factors, the amount of phosphate with 6 levels and the amount of liming with 3 levels. The design was replicated in 3 blocks. The data is input and the analysis of variance computed. The analysis of variance table and tables of means with their standard errors are printed.

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.

```
*     G04CAF Example Program Text
*     Mark 16 Release. NAG Copyright 1992.
*     .. Parameters ..
      INTEGER          NIN, NOUT
      PARAMETER        (NIN=5,NOUT=6)
      INTEGER          NMAX, MAXF, MAXT, MTERM, BMAX
      PARAMETER        (NMAX=54,MAXF=2,MAXT=27,MTERM=6,BMAX=4)
*     .. Local Scalars ..
      INTEGER          I, IFAIL, INTER, IRDF, ITOTAL, J, K, L, N,
     +                 NBLOCK, NFAC, NTREAT
*     .. Local Arrays ..
      real             BMEAN(BMAX), E(MAXT), R(NMAX), SEMEAN(MTERM),
     +                 TABLE(MTERM,5), TMEAN(MAXT), Y(NMAX)
      INTEGER          IMEAN(MTERM), IWK(NMAX+3*MAXF), LFAC(MAXF)
*     .. External Subroutines ..
      EXTERNAL         G04CAF
*     .. Executable Statements ..
      WRITE (NOUT,FMT=*) 'G04CAF Example Program Results'
*     Skip heading in data file
      READ (NIN,FMT=*)
      READ (NIN,FMT=*) N, NBLOCK, NFAC, INTER
      IF (N.LE.NMAX .AND. NBLOCK.LE.BMAX-1 .AND. NFAC.LE.MAXF) THEN
         READ (NIN,FMT=*) (LFAC(J),J=1,NFAC)
         READ (NIN,FMT=*) (Y(I),I=1,N)
         IRDF = 0
         IFAIL = -1
*
         CALL G04CAF(N,Y,NFAC,LFAC,NBLOCK,INTER,IRDF,MTERM,TABLE,ITOTAL,
     +               TMEAN,MAXT,E,IMEAN,SEMEAN,BMEAN,R,IWK,IFAIL)
*
         WRITE (NOUT,FMT=*)
         WRITE (NOUT,FMT=*) ' ANOVA table'
         WRITE (NOUT,FMT=*)
         WRITE (NOUT,FMT=*)
     +     '  Source      df         SS          MS          F',
     +     '        Prob'
         WRITE (NOUT,FMT=*)
         K = 0
         IF (NBLOCK.GT.1) THEN
            K = K + 1
            WRITE (NOUT,FMT=99998) ' Blocks    ', (TABLE(1,J),J=1,5)
         END IF
         NTREAT = ITOTAL - 2 - K
         DO 20 I = 1, NTREAT
            WRITE (NOUT,FMT=99997) ' Effect  ', I, (TABLE(K+I,J),J=1,5)
```

```
   20    CONTINUE
         WRITE (NOUT,FMT=99998) ' Residual  ', (TABLE(ITOTAL-1,J),J=1,3)
         WRITE (NOUT,FMT=99998) ' Total     ', (TABLE(ITOTAL,J),J=1,2)
         WRITE (NOUT,FMT=*)
         WRITE (NOUT,FMT=*) ' Treatment Means and Standard Errors'
         WRITE (NOUT,FMT=*)
         K = 1
         DO 40 I = 1, NTREAT
            L = IMEAN(I)
            WRITE (NOUT,FMT=99996) ' Effect ', I
            WRITE (NOUT,FMT=*)
            WRITE (NOUT,FMT=99999) (TMEAN(J),J=K,L)
            WRITE (NOUT,FMT=*)
            WRITE (NOUT,FMT=99995) ' SE of difference in means  = ',
     +        SEMEAN(I)
            WRITE (NOUT,FMT=*)
            K = L + 1
   40    CONTINUE
      END IF
      STOP
*
99999 FORMAT (8F10.2)
99998 FORMAT (A,3X,F3.0,2X,2(F10.0,2X),F10.3,2X,F9.4)
99997 FORMAT (A,I2,3X,F3.0,2X,2(F10.0,2X),F10.3,2X,F9.4)
99996 FORMAT (A,I2)
99995 FORMAT (A,F10.2)
      END
```

9.2 Program Data

```
G04CAF Example Program Data
54 3 2 2  : N NBLOCK NFAC INTER
6 3       : LFAC

274 361 253 325 317 339 326 402 336 379 345 361 352 334 318 339 393 358
350 340 203 397 356 298 382 376 355 418 387 379 432 339 293 322 417 342
 82 297 133 306 352 361 220 333 270 388 379 274 336 307 266 389 333 353
```

9.3 Program Results

```
 G04CAF Example Program Results

  ANOVA table

   Source      df          SS           MS           F         Prob

 Blocks         2.      30119.      15059.       7.685     0.0018
 Effect   1     5.      73008.      14602.       7.451     0.0001
 Effect   2     2.      21596.      10798.       5.510     0.0085
 Effect   3    10.      31192.       3119.       1.592     0.1513
 Residual      34.      66628.       1960.
 Total         53.     222543.

  Treatment Means and Standard Errors

 Effect  1

    254.78    339.00    333.33    367.78    330.78    360.67

 SE of difference in means  =      20.87

 Effect  2

    334.28    353.78    305.11

 SE of difference in means  =      14.76

 Effect  3

    235.33    332.67    196.33    342.67    341.67    332.67    309.33    370.33
```

```
   320.33    395.00    370.33    338.00    373.33    326.67    292.33    350.00
   381.00    351.00

 SE of difference in means  =      36.14
```