NAG Fortran Library Routine Document

G04BBF

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

G04BBF computes the analysis of variance and treatment means and standard errors for a randomized block or completely randomized design.

2 Specification

```
 SUBROUTINE G04BBF(N, Y, IBLOCK, NT, IT, GMEAN, BMEAN, TMEAN, TABLE, LDT,
1                  C, LDC, IREP, R, EF, TOL, IRDF, WK, IFAIL)
 INTEGER           N, IBLOCK, NT, IT(*), LDT, LDC, IREP(NT), IRDF, IFAIL
 real              Y(N), GMEAN, BMEAN(*), TMEAN(NT), TABLE(LDT,5),
1                  C(LDC,NT), R(N), EF(NT), TOL, WK(NT*NT+NT)
```

3 Description

In a completely randomized design the experimental material is divided into a number of units, or plots, to which a treatment can be applied. In a randomized block design the units are grouped into blocks so that the variation within blocks is less than the variation between blocks. If every treatment is applied to one plot in each block it is a complete block design. If there are fewer plots per block than treatments then the design will be an incomplete block design and may be balanced or partially balanced.

For a completely randomized design, with t treatments and n_t plots per treatment, the linear model is:

$$y_{ij} = \mu + \tau_j + e_{ij}, \quad j = 1, 2, \ldots, t; \quad i = 1, 2, \ldots, n_j,$$

where y_{ij} is the ith observation for the jth treatment, μ is the overall mean, τ_j is the effect of the jth treatment and e_{ij} is the random error term. For a randomised block design, with t treatments and b blocks of k plots, the linear model is:

$$y_{ij(l)} = \mu + \beta_i + \tau_l + e_{ij}, \quad i = 1, 2, \ldots, b; \quad j = 1, 2, \ldots, k; \quad l = 1, 2, \ldots, t,$$

where β_i is the effect of the ith block and the $ij(l)$ notation indicates that the lth treatment is applied to the ith plot in the jth block.

The completely randomized design gives rise to a one-way analysis of variance. The treatments do not have to be equally replicated, i.e., do not have to occur the same number of times. First the overall mean, $\hat{\mu}$, is computed and subtracted from the observations to give, $y'_{ij} = y_{ij} - \hat{\mu}$. The estimated treatment effects, $\hat{\tau}_j$ are then computed as the treatment means of the mean adjusted observations, y'_{ij}, and the treatment sum of squares can be computed from the sum of squares of the treatment totals of the y'_{ij} divided by the number of observations per treatment total, n_j. The final residuals are computed as $r_{ij} = y'_{ij} - \hat{\tau}_j$, and, from the residuals, the residual sum of squares is calculated.

For the randomised block design the mean is computed and subtracted from the observations to give $y'_{ij(l)} = y_{ij(l)} - \hat{\mu}$. The estimated block effects, ignoring treatment effects, $\hat{\beta}_i$, are then computed using the block means of the $y'_{ij(l)}$ and the unadjusted sum of squares computed as the sum of squared block totals for the $y'_{ij(l)}$ divided by number of plots per block, k. The block adjusted observations are then computed as $y''_{ij(l)} = y'_i j_{(l)} = \hat{\beta}_i$. In the case of the complete block design, with the same replication for each treatment within each block, the blocks and treatments are orthogonal, and so the treatment effects are estimated as the treatment means of the block adjusted observations, $y''_{ij(l)}$. The treatment sum of squares is computed as the sum of squared treatment totals of the $y''_{ij(l)}$ divided by the number of replicates to the

treatments, $r = bk/t$. Finally the residuals, and hence the residual sum of squares, are given by $r_{ij(l)} = y''_{ij(l)} - \hat{\tau}_l$.

For a design without the same replication for each treatment within each block the treatments and the blocks will not be orthogonal, so the treatments adjusted for blocks need to be computed. The adjusted treatment effects are found as the solution to the equations:

$$(R - NN^{\mathrm{T}}/k)\hat{\tau} = q$$

where q is the vector of the treatment totals for block adjusted observations, $y''_{ij(l)}$, R is a diagonal matrix with R_{ll} equal to the number of times the lth treatment is replicated, and N is the t by b incidence matrix, with N_{lj} equal to the number of times treatment l occurs in block j. The solution to the equations can be written as:

$$\hat{\tau} = \Omega q$$

where Ω is a generalized inverse of $(R - NN^{\mathrm{T}}/k)$. The solution is found from the eigenvalue decomposition of $(R - NN^{\mathrm{T}}/k)$. The residuals are first calculated by subtracting the estimated treatment effects from the block adjusted observations to give $r'_{ij(l)} = y''_{ij(l)} - \hat{\tau}_l$. However, since only the unadjusted block effects have been removed and blocks and treatments are not orthogonal, the block means of the $r'_{ij(l)}$ have to give the correct residuals, $r_{ij(l)}$ and residual sum of squares.

The mean squares are computed as the sum of squares divided by the degrees of freedom. The degrees of freedom for the unadjusted blocks is $b - 1$, for the completely randomised and the complete block designs the degrees of freedom for the treatments is $t - 1$. In the general case the degrees of freedom for treatments is the rank of the matrix Ω. The F-statistic given by the ratio of the treatment mean square to the residual mean square tests the hypothesis:

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_t = 0.$$

The standard errors for the difference in treatment effects, or treatment means, for the completely randomized or the complete block designs, are given by:

$$se(\tau_j - \tau_{j*}) = \left(\frac{1}{n_j} + \frac{1}{n_{j*}}\right)s^2$$

where s^2 is the residual mean square and $n_j = n_{j*} = b$ in the complete block design. In the general case the variances of the treatment effects are given by:

$$\mathrm{var}(\tau) = \Omega s^2$$

from which the appropriate standard errors of the difference between treatment effects or the difference between adjusted means can be calculated.

In the complete block design all the information on the treatment effects is given by the within block analysis. In other designs there may be a loss of information due to the non-orthogonality of treatments and blocks. The efficiency of the within block analysis in these cases is given by the (canonical) efficiency factors, these are the non-zero eigenvalues of the matrix $(R - NN^{\mathrm{T}}/k)$, divided by the number of replicates in the case of equal replication, or by the mean of the number of replicates in the unequally replicated case, see John (1987). If more than one eigenvalue is zero then the design is said to be disconnected and some treatments can only be compared using a between block analysis.

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 (1987) *Cyclic Designs* Chapman and Hall

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

Searle S R (1971) *Linear Models* Wiley

5 Parameters

1: N – INTEGER *Input*

On entry: the number of observations.

Constraints: N ≥ 2 and if abs(IBLOCK) ≥ 2, N must be a multiple of abs(IBLOCK).

2: Y(N) – ***real*** array *Input*

On entry: the observations in the order as described by IBLOCK and NT.

3: IBLOCK – INTEGER *Input*

On entry: IBLOCK indicates the block structure. If abs(IBLOCK) ≤ 1 then there are no blocks, i.e., it is a completely randomized design. If IBLOCK ≥ 2 then there are IBLOCK blocks and the data should be input by blocks, i.e., Y must contain the observations for block 1 followed by the observations for block 2 etc.. If IBLOCK ≤ -2 then there are abs(IBLOCK) blocks and the data is input in parallel with respect to blocks, i.e., Y(1) must contain the first observation for block 1, Y(2) must contain the first observation for block $2 \cdots$ Y(abs(IBLOCK)) must contain the first observation for block abs(IBLOCK), Y(abs(IBLOCK + 1)) must contain the second observation for block 1, etc.

4: NT – INTEGER *Input*

On entry: the number of treatments. If only blocks are required in the analysis then set NT=1.

Constraint: if abs(IBLOCK) ≥ 2, NT ≥ 1, otherwise NT ≥ 2.

5: IT(∗) – INTEGER array *Input*

Note: the dimension of the array IT must be at least N if NT ≥ 2, and 1 otherwise.

On entry: IT(i) indicates which of the NT treatments plot i received, for $i = 1, 2, \ldots,$ N. If NT=1, IT is not referenced.

Constraint: $1 \leq$ IT(i) $\leq$ NT , for $i = 1, 2, \ldots,$ N.

6: GMEAN – ***real*** *Output*

On exit: the grand mean, $\hat{\mu}$.

7: BMEAN(∗) – ***real*** array *Output*

Note: the dimension of the array BMEAN=1,abs("x(82)=1,abs(list(82) at least max(1,abs(IBLOCK)).

On exit: if abs(IBLOCK) ≥ 2, BMEAN(j) contains the mean for the jth block, $\hat{\beta}_j$ for $j = 1, 2, \ldots, b$.

8: TMEAN(NT) – ***real*** array *Output*

On exit: if NT ≥ 2, TMEAN(l) contains the (adjusted) mean for the lth treatment, $\hat{\mu}^* + \hat{\tau}_l$, for $l = 1, 2, \ldots, t$, where $\hat{\mu}^*$ is the mean of the treatment adjusted observations, $y_{ij(l)} - \hat{\tau}_l$.

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

On exit: the analysis of variance table. Column 1 contains the degrees of freedom, column 2 the sum of squares, and where appropriate, column 3 the mean squares, column 4 the F-statistic and column 5 the significance level of the F-statistic. Row 1 is for Blocks, row 2 for Treatments, row 3 for Residual and row 4 for Total. Mean squares are computed for all but the Total row; F-statistics and significance are computed for Treatments and Blocks, if present. Any unfilled cells are set to zero.

10: LDT – INTEGER *Input*

On entry: the first dimension of the array TABLE as declared in the (sub)program from which G04BBF is called.

Constraint: LDT $\geq$ 4.

11: C(LDC,NT) – ***real*** array *Output*

On exit: if NT $\geq$ 2, the upper triangular part of C contains the variance-covariance matrix of the treatment effects, the strictly lower triangular part contains the standard errors of the difference between two treatment effects (means), i.e., C(i,j) contains the covariance of treatment i and j if $j \geq i$ and the standard error of the difference between treatment i and j if $j < i$ for $i = 1,2,\ldots,t$; $j = 1,2,\ldots,t$.

12: LDC – INTEGER *Input*

On entry: the first dimension of the array C as declared in the (sub)program from which G04BBF is called.

Constraint: LDC $\geq$ NT.

13: IREP(NT) – INTEGER array *Output*

On exit: if NT $\geq$ 2, the treatment replications, R_{ll}, for $l = 1,2,\ldots,$NT.

14: R(N) – ***real*** array *Output*

On exit: the residuals, r_i, for $i = 1,2,\ldots$N.

15: EF(NT) – ***real*** array *Output*

On exit: if NT $\geq$ 2, the canonical efficiency factors.

16: TOL – ***real*** *Input*

On entry: the tolerance value used to check for zero eigenvalues of the matrix Ω. If TOL=0.0 a default value of 10^{-5} is used.

Constraint: TOL $\geq$ 0.0.

17: IRDF – INTEGER *Input*

On entry: an adjustment to the degrees of freedom for the residual and total. If IRDF $\geq$ 1 the degrees of freedom for the total is set to N $-$ IRDF and the residual degrees of freedom adjusted accordingly. If IRDF=0, the total degrees of freedom for the total is set to N $-$ 1, as usual.

Constraint: IRDF $\geq$ 0.

18: WK(NT*NT+NT) – ***real*** 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, 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, N < 2,
or NT ≤ 0,
or NT $= 1$ and abs(IBLOCK) ≤ 1,
or LDT < 4,
or LDC $<$ NT,
or TOL < 0.0,
or IRDF < 0.

IFAIL $= 2$

On entry, abs(IBLOCK) ≥ 2 and N is not a multiple of abs(IBLOCK).

IFAIL $= 3$

On entry, IT$(i) < 1$ or IT$(i) >$ NT for some i when NT ≥ 2,
or no value of IT $= j$ for some $j = 1, 2, \ldots,$ NT, when NT ≥ 2.

IFAIL $= 4$

On entry, the values of Y are constant.

IFAIL $= 5$

A computed standard error is zero due to rounding errors, or the eigenvalue computation failed to converge. Both are unlikely error exits.

IFAIL $= 6$

The treatments are totally confounded with blocks, so the treatment sum of squares and degrees of freedom are zero. The analysis of variance table is not computed, except for block and total sums of squares and degrees of freedom.

IFAIL $= 7$

The residual degrees of freedom or the residual sum of squares are zero, columns 3, 4 and 5 of the analysis of variance table will not be computed and the matrix of standard errors and covariances, C, will not be scaled by s or s^2.

IFAIL $= 8$

The design is disconnected; the standard errors may not be valid. The design may be nested.

7 Accuracy

The algorithm used by this routine, described in Section 3, achieves greater accuracy than the traditional algorithms based on the subtraction of sums of squares.

8 Further Comments

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 $\hat{\mu}$.

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 designs such as Graeco-Latin squares one or more of the blocking factors has to be removed in a preliminary analysis before the final analysis using calls to G04BBF or G04BCF. The residuals from the preliminary analysis are then input to G04BBF. 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.

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 results from those routines can be used to test for the significance of the covariates. To test the significance of the treatment effects after fitting the covariate, the residual sum of squares from the regression should be compared with the residual sum of squares obtained from the equivalent regression but using the residuals from fitting blocks only.

9 Example

The data, given by John and Quenouille (1977), are for a balanced incomplete block design with 10 blocks and 6 treatments and with 3 plots per block. The observations are the degree of pain experienced and the treatments are penicillin of different potency. The data is input and the analysis of variance table and treatment means 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.

```
*     G04BBF Example Program Text
*     Mark 16 Release. NAG Copyright 1992.
*     .. Parameters ..
      INTEGER          NIN, NOUT
      PARAMETER        (NIN=5,NOUT=6)
      INTEGER          NMAX, NTMAX, NBMAX, TMAX
      PARAMETER        (NMAX=30,NTMAX=6,NBMAX=10,TMAX=4)
*     .. Local Scalars ..
      real             GMEAN, TOL
      INTEGER          I, IFAIL, IRDF, J, N, NBLOCK, NT
*     .. Local Arrays ..
      real             BMEAN(NBMAX), C(NTMAX,NTMAX), EF(NTMAX), R(NMAX),
     +                 TABLE(TMAX,5), TMEAN(NTMAX),
     +                 WK(NTMAX*NTMAX+NTMAX), Y(NMAX)
      INTEGER          IREP(NTMAX), IT(NMAX)
*     .. External Subroutines ..
      EXTERNAL         G04BBF
*     .. Executable Statements ..
      WRITE (NOUT,FMT=*) 'G04BBF Example Program Results'
*     Skip heading in data file
      READ (NIN,FMT=*)
      READ (NIN,FMT=*) N, NT, NBLOCK
      IF (N.LE.NMAX) THEN
         READ (NIN,FMT=*) (Y(I),I=1,N)
         READ (NIN,FMT=*) (IT(I),I=1,N)
         TOL = 0.000005e0
         IRDF = 0
         IFAIL = -1
*
         CALL G04BBF(N,Y,NBLOCK,NT,IT,GMEAN,BMEAN,TMEAN,TABLE,TMAX,C,
     +              NTMAX,IREP,R,EF,TOL,IRDF,WK,IFAIL)
*
         WRITE (NOUT,FMT=*)
         WRITE (NOUT,FMT=*) ' ANOVA table'
         WRITE (NOUT,FMT=*)
         WRITE (NOUT,FMT=*)
     +     '  Source        df         SS          MS          F',
     +     '        Prob'
```

```
         WRITE (NOUT,FMT=*)
         WRITE (NOUT,FMT=99998) ' Blocks      ', (TABLE(1,J),J=1,5)
         WRITE (NOUT,FMT=99998) ' Treatments  ', (TABLE(2,J),J=1,5)
         WRITE (NOUT,FMT=99998) ' Residual    ', (TABLE(3,J),J=1,3)
         WRITE (NOUT,FMT=99998) ' Total       ', (TABLE(4,J),J=1,2)
         WRITE (NOUT,FMT=*)
         WRITE (NOUT,FMT=*) ' Efficiency Factors'
         WRITE (NOUT,FMT=*)
         WRITE (NOUT,FMT=99999) (EF(I),I=1,NT)
         WRITE (NOUT,FMT=*)
         WRITE (NOUT,FMT=99997) '  Grand Mean', GMEAN
         WRITE (NOUT,FMT=*)
         WRITE (NOUT,FMT=*) ' Treatment Means'
         WRITE (NOUT,FMT=*)
         WRITE (NOUT,FMT=99999) (TMEAN(I),I=1,NT)
         WRITE (NOUT,FMT=*)
         WRITE (NOUT,FMT=*)
     +     ' Standard errors of differences between means'
         WRITE (NOUT,FMT=*)
         DO 20 I = 2, NT
            WRITE (NOUT,FMT=99999) (C(I,J),J=1,I-1)
   20    CONTINUE
      END IF
      STOP
*
99999 FORMAT (8F10.2)
99998 FORMAT (A,3X,F3.0,2X,3(F10.2,2X),F9.4)
99997 FORMAT (A,F10.2)
      END
```

9.2 Program Data

```
G04BBF Example Program Data
30 6 10                     : N, NT, IBLOCK
1 5 4
5 10 6
2 9 3
4 8 6
2 4 7
6 7 5
5 7 2
7 2 4
8 4 2
10 8 7                      : Y
1 2 3
1 2 4
1 3 5
1 4 6
1 5 6
2 3 6
2 4 5
2 5 6
3 4 5
3 4 6                       : IT
```

9.3 Program Results

```
 G04BBF Example Program Results

  ANOVA table

   Source        df         SS          MS          F        Prob

 Blocks           9.       60.00        6.67        4.79     0.0039
 Treatments       5.      101.78       20.36       14.62     0.0000
 Residual        15.       20.89        1.39
 Total           29.      182.67

  Efficiency Factors

      0.00      0.80      0.80      0.80      0.80      0.80

  Grand Mean      5.33

  Treatment Means

      2.50      7.25      8.08      5.92      2.92      5.33

  Standard errors of differences between means

      0.83
      0.83      0.83
      0.83      0.83      0.83
      0.83      0.83      0.83      0.83
      0.83      0.83      0.83      0.83      0.83
```