NAG Library Routine Document

G04AGF

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

G04AGF performs an analysis of variance for a two-way hierarchical classification with subgroups of possibly unequal size, and also computes the treatment group and subgroup means. A fixed effects model is assumed.

2 Specification

```
SUBROUTINE G04AGF (Y, N, K, LSUB, NOBS, L, NGP, GBAR, SGBAR, GM, SS, IDF,       &
                   F, FP, IFAIL)

INTEGER            N, K, LSUB(K), NOBS(L), L, NGP(K), IDF(4), IFAIL
REAL (KIND=nag_wp) Y(N), GBAR(K), SGBAR(L), GM, SS(4), F(2), FP(2)
```

3 Description

In a two-way hierarchical classification, there are k (≥ 2) treatment groups, the ith of which is subdivided into l_i treatment subgroups. The jth subgroup of group i contains n_{ij} observations, which may be denoted by

$$y_{1ij}, y_{2ij}, \ldots, y_{n_{ij}ij}.$$

The general observation is denoted by y_{mij}, being the mth observation in subgroup j of group i, for $1 \leq i \leq k$, $1 \leq j \leq l_i$, $1 \leq m \leq n_{ij}$.

The following quantities are computed

(i) The subgroup means

$$\bar{y}_{.ij} = \frac{\sum_{m=1}^{n_{ij}} y_{mij}}{n_{ij}}$$

(ii) The group means

$$\bar{y}_{.i.} = \frac{\sum_{j=1}^{l_i}\sum_{m=1}^{n_{ij}} y_{mij}}{\sum_{j=1}^{l_i} n_{ij}}$$

(iii) The grand mean

$$\bar{y}_{...} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{l_i}\sum_{m=1}^{n_{ij}} y_{mij}}{\sum_{i=1}^{k}\sum_{j=1}^{l_i} n_{ij}}$$

(iv) The number of observations in each group

$$n_{i.} = \sum_{j=1}^{l_i} n_{ij}$$

(v) Sums of squares

Between groups $= \mathrm{SS}_g = \sum_{i=1}^{k} n_{i.} (\bar{y}_{.i.} - \bar{y}_{...})^2$

Between subgroups within groups $= \mathrm{SS}_{sg} = \sum_{i=1}^{k} \sum_{j=1}^{l_i} n_{ij} (y_{.ij} - \bar{y}_{.i.})^2$

Residual (within subgroups) $= \mathrm{SS}_{\mathrm{res}} = \sum_{i=1}^{k} \sum_{j=1}^{l_i} \sum_{m=1}^{n_{ij}} (y_{mij} - \bar{y}_{.ij})^2 = \mathrm{SS}_{\mathrm{tot}} - \mathrm{SS}_g - \mathrm{SS}_{sg}$

Corrected total $= \mathrm{SS}_{\mathrm{tot}} = \sum_{i=1}^{k} \sum_{j=1}^{l_i} \sum_{m=1}^{n_{ij}} (y_{mij} - \bar{y}_{...})^2$

(vi) Degrees of freedom of variance components

Between groups: $k - 1$
Subgroups within groups: $l - k$
Residual: $n - l$
Total: $n - 1$

where

$$l = \sum_{i=1}^{k} l_i,$$

$$n = \sum_{i=1}^{k} n_{i.}$$

(vii) F ratios. These are the ratios of the group and subgroup mean squares to the residual mean square.

Groups $F_1 = \dfrac{\text{Between groups sum of squares}/(k-1)}{\text{Residual sum of squares}/(n-l)} = \dfrac{\mathrm{SS}_g/(k-1)}{\mathrm{SS}_{\mathrm{res}}/(n-l)}$

Subgroups $F_2 = \dfrac{\text{Between subgroups (within group) sum of squares}/(l-k)}{\text{Residual sum of squares}/(n-l)} = \dfrac{\mathrm{SS}_{sg}/(l-k)}{\mathrm{SS}_{\mathrm{res}}/(n-l)}$

If either F ratio exceeds 9999.0, the value 9999.0 is assigned instead.

(viii) F significances. The probability of obtaining a value from the appropriate F-distribution which exceeds the computed mean square ratio.

Groups $p_1 = \mathrm{Prob}\big(F_{(k-1),(n-l)} > F_1\big)$

Subgroups $p_2 = \mathrm{Prob}\big(F_{(l-k),(n-l)} > F_2\big)$

where F_{ν_1,ν_2} denotes the central F-distribution with degrees of freedom ν_1 and ν_2.

If any $F_i = 9999.0$, then p_i is set to zero, $i = 1, 2$.

4 References

Kendall M G and Stuart A (1976) *The Advanced Theory of Statistics (Volume 3)* (3rd Edition) Griffin

Moore P G, Shirley E A and Edwards D E (1972) *Standard Statistical Calculations* Pitman

5 Parameters

1: Y(N) – REAL (KIND=nag_wp) array *Input*

On entry: the elements of Y must contain the observations y_{mij} in the following order:

$$y_{111}, y_{211}, \ldots, y_{n_{11}11}, y_{112}, y_{212}, \ldots, y_{n_{12}12}, \ldots, y_{11l_1}, \ldots,$$

$$y_{n_{1l_1}1l_1}, \ldots, y_{1ij}, \ldots, y_{n_{ij}ij}, \ldots, y_{1kl_k}, \ldots, y_{n_{kl_k}kl_k}.$$

In words, the ordering is by group, and within each group is by subgroup, the members of each subgroup being in consecutive locations in Y.

2: N – INTEGER *Input*

On entry: n, the total number of observations.

3: K – INTEGER *Input*

On entry: k, the number of groups.

Constraint: $\mathrm{K} \geq 2$.

4: LSUB(K) – INTEGER array *Input*

On entry: the number of subgroups within group i, l_i, for $i = 1, 2, \ldots, k$.

Constraint: $\mathrm{LSUB}(i) > 0$, for $i = 1, 2, \ldots, k$.

5: NOBS(L) – INTEGER array *Input*

On entry: the numbers of observations in each subgroup, n_{ij}, in the following order:

$$n_{11}, n_{12}, \ldots, n_{1l_1}, n_{21}, \ldots, n_{2l_2}, \ldots, n_{k1}, \ldots, n_{kl_k}$$

Constraint: $n = \sum_{i=1}^{k} \sum_{j=1}^{l_i} n_{ij}$, that is $\mathrm{N} = \sum_{i=1}^{l} \mathrm{NOBS}(i)$ and $\mathrm{NOBS}(i) > 0$, for $i = 1, 2, \ldots, l$.

6: L – INTEGER *Input*

On entry: l, the total number of subgroups.

Constraint: $\mathrm{L} = \sum_{i=1}^{k} \mathrm{LSUB}(i)$.

7: NGP(K) – INTEGER array *Output*

On exit: the total number of observations in group i, $n_{i.}$, for $i = 1, 2, \ldots, k$.

8: GBAR(K) – REAL (KIND=nag_wp) array *Output*

On exit: the mean for group i, $\bar{y}_{.i.}$, for $i = 1, 2, \ldots, k$.

9: SGBAR(L) – REAL (KIND=nag_wp) array *Output*

On exit: the subgroup means, $\bar{y}_{.ij}$, in the following order:

$$\bar{y}_{.11}, \bar{y}_{.12}, \ldots, \bar{y}_{.1l_1}, \bar{y}_{.21}, \bar{y}_{.22}, \ldots, \bar{y}_{.2l_2}, \ldots, \bar{y}_{.k1}, \bar{y}_{.k2}, \ldots, \bar{y}_{.kl_k}.$$

10: GM – REAL (KIND=nag_wp) *Output*

On exit: the grand mean, $\bar{y}_{...}$.

11: SS(4) – REAL (KIND=nag_wp) array *Output*

On exit: contains the sums of squares for the analysis of variance, as follows;

SS(1) = Between group sum of squares, SS_g,

SS(2) = Between subgroup within groups sum of squares, SS_{sg},

SS(3) = Residual sum of squares, SS_{res},

SS(4) = Corrected total sum of squares, SS_{tot}.

12: IDF(4) – INTEGER array *Output*

On exit: contains the degrees of freedom attributable to each sum of squares in the analysis of variance, as follows:

IDF(1) = Degrees of freedom for between group sum of squares,

IDF(2) = Degrees of freedom for between subgroup within groups sum of squares,

IDF(3) = Degrees of freedom for residual sum of squares,

IDF(4) = Degrees of freedom for corrected total sum of squares.

13: F(2) – REAL (KIND=nag_wp) array *Output*

On exit: contains the mean square ratios, F_1 and F_2, for the between groups variation, and the between subgroups within groups variation, with respect to the residual, respectively.

14: FP(2) – REAL (KIND=nag_wp) array *Output*

On exit: contains the significances of the mean square ratios, p_1 and p_2 respectively.

15: 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, K ≤ 1.

IFAIL $= 2$

On entry, LSUB$(i) \leq 0$, for some $i = 1, 2, \ldots, k$.

IFAIL $= 3$

On entry, L $\neq \sum_{i=1}^{k}$LSUB(i)

IFAIL $= 4$

On entry, NOBS$(i) \leq 0$, for some $i = 1, 2, \ldots, l$.

IFAIL = 5

On entry, $N \neq \sum_{i=1}^{l} \text{NOBS}(i)$.

IFAIL = 6

The total corrected sum of squares is zero, indicating that all the data values are equal. The means returned are therefore all equal, and the sums of squares are zero. No assignments are made to IDF, F, and FP.

IFAIL = 7

The residual sum of squares is zero. This arises when either each subgroup contains exactly one observation, or the observations within each subgroup are equal. The means, sums of squares, and degrees of freedom are computed, but no assignments are made to F and FP.

7 Accuracy

The computations are believed to be stable.

8 Further Comments

The time taken by G04AGF increases approximately linearly with the total number of observations, n.

9 Example

This example has two groups, the first of which consists of five subgroups, and the second of three subgroups. The numbers of observations in each subgroup are not equal. The data represent the percentage stretch in the length of samples of sack kraft drawn from consignments (subgroups) received over two years (groups). For details see Moore *et al.* (1972).

9.1 Program Text

```
    Program g04agfe

!     G04AGF Example Program Text

!     Mark 24 Release. NAG Copyright 2012.

!     .. Use Statements ..
      Use nag_library, Only: g04agf, nag_wp
!     .. Implicit None Statement ..
      Implicit None
!     .. Parameters ..
      Integer, Parameter              :: nin = 5, nout = 6
!     .. Local Scalars ..
      Real (Kind=nag_wp)              :: gm
      Integer                         :: i, ifail, ii, j, k, l, li, n, nhi,   &
                                         nij, nlo, nsub
!     .. Local Arrays ..
      Real (Kind=nag_wp)              :: f(2), fp(2), ss(4)
      Real (Kind=nag_wp), Allocatable :: gbar(:), sgbar(:), y(:)
      Integer                         :: idf(4)
      Integer, Allocatable            :: lsub(:), ngp(:), nobs(:)
!     .. Intrinsic Procedures ..
      Intrinsic                       :: sum
!     .. Executable Statements ..
      Write (nout,*) 'G04AGF Example Program Results'
      Write (nout,*)

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

```
!     Read in number of groups
      Read (nin,*) k

      Allocate (lsub(k),ngp(k),gbar(k))

!     Read in number of subgroups
      Read (nin,*) lsub(1:k)

!     Total number of subgroups
      l = sum(lsub(1:k))

      Allocate (nobs(l),sgbar(l))

!     Read in the number of observations
      Read (nin,*) nobs(1:l)

!     Total number of observations
      n = sum(nobs(1:l))

      Allocate (y(n))

!     Read in the data
      Read (nin,*) y(1:n)

!     Display data
      Write (nout,*) 'Data values'
      Write (nout,*)
      Write (nout,*) ' Group  Subgroup  Observations'
      nsub = 0
      nlo = 1
      Do i = 1, k
        li = lsub(i)
        Do j = 1, li
          nsub = nsub + 1
          nij = nobs(nsub)
          nhi = nlo + nij - 1
          Write (nout,99999) i, j, y(nlo:nhi)
          nlo = nlo + nij
        End Do
      End Do

!     Perform ANOVA
      ifail = 0
      Call g04agf(y,n,k,lsub,nobs,l,ngp,gbar,sgbar,gm,ss,idf,f,fp,ifail)

!     Display results
      Write (nout,*)
      Write (nout,*) 'Subgroup means'
      Write (nout,*)
      Write (nout,*) '   Group  Subgroup  Mean'
      ii = 0
      Do i = 1, k
        li = lsub(i)
        Do j = 1, li
          ii = ii + 1
          Write (nout,99998) i, j, sgbar(ii)
        End Do
      End Do
      Write (nout,*)
      Write (nout,99997) '    Group 1 mean =', gbar(1), '   (', ngp(1), &
        ' observations)'
      Write (nout,99997) '    Group 2 mean =', gbar(2), '   (', ngp(2), &
        ' observations)'
      Write (nout,99997) '    Grand mean   =', gm, '   (', n, ' observations)'
      Write (nout,*)
      Write (nout,*) 'Analysis of variance table'
      Write (nout,*)
      Write (nout,*) '   Source                SS    DF  F ratio  Sig'
      Write (nout,*)
      Write (nout,99996) 'Between groups        ', ss(1), idf(1), f(1), fp(1)
      Write (nout,99996) 'Bet sbgps within gps  ', ss(2), idf(2), f(2), fp(2)
```

```
      Write (nout,99996) 'Residual              ', ss(3), idf(3)
      Write (nout,*)
      Write (nout,99996) 'Total                 ', ss(4), idf(4)

99999 Format (1X,I5,I9,4X,10F4.1)
99998 Format (1X,I6,I8,F10.2)
99997 Format (1X,A,F5.2,A,I2,A)
99996 Format (1X,A,F6.3,I5,F7.2,F8.3)
    End Program g04agfe
```

9.2 Program Data

```
G04AGF Example Program Data
 2                             :: K (number of groups)
 5 3                           :: LSUB (number of subgroups per group)
 5 3 3 3 2 3 5 3               :: NOBS (number of observations per subgroup)
 2.1 2.4 2.0 2.0 2.0
 2.4 2.1 2.2 2.4 2.2
 2.6 2.4 2.4 2.5 1.9
 1.7 2.1 1.5 2.0 1.9
 1.7 1.9 1.9 1.9 2.0 2.1 2.3   :: Y (observed data)
```

9.3 Program Results

```
 G04AGF Example Program Results

 Data values

  Group  Subgroup  Observations
     1       1      2.1 2.4 2.0 2.0 2.0
     1       2      2.4 2.1 2.2
     1       3      2.4 2.2 2.6
     1       4      2.4 2.4 2.5
     1       5      1.9 1.7
     2       1      2.1 1.5 2.0
     2       2      1.9 1.7 1.9 1.9 1.9
     2       3      2.0 2.1 2.3

 Subgroup means

    Group  Subgroup  Mean
      1       1      2.10
      1       2      2.23
      1       3      2.40
      1       4      2.43
      1       5      1.80
      2       1      1.87
      2       2      1.86
      2       3      2.13

     Group 1 mean = 2.21   (16 observations)
     Group 2 mean = 1.94   (11 observations)
     Grand mean   = 2.10   (27 observations)

 Analysis of variance table

    Source                 SS    DF  F ratio  Sig

 Between groups         0.475    1  16.15   0.001
 Bet sbgps within gps   0.816    6   4.63   0.005
 Residual               0.559   19

 Total                  1.850   26
```