NAG Toolbox for MATLAB

g04ag

1 Purpose

g04ag 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 Syntax

```
[ngp, gbar, sgbar, gm, ss, idf, f, fp, ifail] = g04ag(y, lsub, nobs,
'n', n, 'k', k, 'l', l)
```

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

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

(ii) The group means

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

(iii) The grand mean

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

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

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

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

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

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

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

5.1 Compulsory Input Parameters

1: **y(n) – double array**

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: **lsub(k) – int32 array**

The number of subgroups within group i, l_i, for $i = 1, 2, \ldots, k$.

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

3: **nobs(l) – int32 array**

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 $\mathbf{n} = \sum_{i=1}^{l} \mathbf{nobs}(i)$ and $\mathbf{nobs}(i) > 0$, for $i = 1, 2, \ldots, l$.

5.2 Optional Input Parameters

1: **n – int32 scalar**

Default: The dimension of the array **y**.

n, the total number of observations.

2: **k – int32 scalar**

Default: The dimension of the arrays **lsub**, **gbar**. (An error is raised if these dimensions are not equal.)

k, the number of groups.

Constraint: $\mathbf{k} \geq 2$.

3: **l – int32 scalar**

Default: The dimension of the arrays **nobs**, **sgbar**. (An error is raised if these dimensions are not equal.)

l, the total number of subgroups.

Constraint: $\mathbf{l} = \sum_{i=1}^{k} \mathbf{lsub}(i)$.

5.3 Input Parameters Omitted from the MATLAB Interface

None.

5.4 Output Parameters

1: **ngp(k) – int32 array**

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

2: **gbar(k) – double array**

The mean for group i, $\bar{y}_{.i.}$, for $i = 1, 2, \ldots, k$.

3: **sgbar(l) – double array**

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

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

4: **gm – double scalar**

The grand mean, $\bar{y}_{...}$.

5: **ss(4) – double array**

Contains the sums of squares for the analysis of variance, as follows;

ss(1) = Between group sum of squares, $\mathbf{ss}_g$,

ss(2) = Between subgroup within groups sum of squares, $\mathbf{ss}_{sg}$,

ss(3) = Residual sum of squares, $\mathbf{ss}_{\mathrm{res}}$,

ss(4) = Corrected total sum of squares, $\mathbf{ss}_{\mathrm{tot}}$.

6: **idf(4) – int32 array**

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.

7: **f(2) – double array**

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.

8: **fp(2) – double array**

Contains the significances of the mean square ratios, p_1 and p_2 respectively.

9: **ifail – int32 scalar**

0 unless the function detects an error (see Section 6).

6 Error Indicators and Warnings

Errors or warnings detected by the function:

ifail = 1

On entry, $\mathbf{k} \leq 1$.

ifail $= 2$

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

ifail $= 3$

On entry, $\mathbf{l} \neq \sum_{i=1}^{k} \mathbf{lsub}(i)$

ifail $= 4$

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

ifail $= 5$

On entry, $\mathbf{n} \neq \sum_{i=1}^{l} \mathbf{nobs}(i)$.

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 g04ag increases approximately linearly with the total number of observations, n.

9 Example

```
y = [2.1;
     2.4;
     2;
     2;
     2;
     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;
     1.9;
     1.7;
```

```
     1.9;
     1.9;
     1.9;
     2;
     2.1;
     2.3];
lsub = [int32(5);
     int32(3)];
nobs = [int32(5);
     int32(3);
     int32(3);
     int32(3);
     int32(2);
     int32(3);
     int32(5);
     int32(3)];
[ngp, gbar, sgbar, gm, ss, idf, f, fp, ifail] = g04ag(y, lsub, nobs)
```

```
ngp =
          16
          11
gbar =
    2.2062
    1.9364
sgbar =
    2.1000
    2.2333
    2.4000
    2.4333
    1.8000
    1.8667
    1.8600
    2.1333
gm =
    2.0963
ss =
    0.4748
    0.8162
    0.5587
    1.8496
idf =
           1
           6
          19
          26
f =
   16.1477
    4.6262
fp =
    0.0007
    0.0047
ifail =
           0
```