NAG Toolbox for MATLAB

g04bc

1   Purpose

g04bc computes the analysis of variance for a general row and column design together with the treatment means and standard errors.

2   Syntax

```
[gmean, tmean, tabl, c, irep, rpmean, rmean, cmean, r, ef, ifail] =
g04bc(nrep, nrow, ncol, y, nt, it, tol, irdf)
```

3   Description

In a row and column design the experimental material can be characterized by a two-way classification, nominally called rows and columns. Each experimental unit can be considered as being located in a particular row and column. It is assumed that all rows are of the same length and all columns are of the same length. Sets of equal numbers of rows/columns can be grouped together to form replicates, sometimes known as squares or rectangles, as appropriate.

If for a replicate, the number of rows, the number of columns and the number of treatments are equal and every treatment occurs once in each row and each column then the design is a Latin square. If this is not the case the treatments will be non-orthogonal to rows and columns. For example in the case of a lattice square each treatment occurs only once in each square.

For a row and column design, with t treatments in r rows and c columns and b replicates or squares with $n = brc$ observations the linear model is:

$$y_{ijk(l)} = \mu + \beta_i + \rho_j + \gamma_k + \tau_l + e_{ijk}$$

for $i = 1, 2 \ldots, b; j = 1, 2, \ldots, r; k = 1, 2 \ldots, c; l = 1, 2, \ldots, t$, where β_i is the effect of the ith replicate, ρ_j is the effect of the jth row, γ_k is the effect of the kth column and the $ijk(l)$ notation indicates that the lth treatment is applied to the unit in row j, column k of replicate i.

To compute the analysis of variance for a row and column design the mean is computed and subtracted from the observations to give, $y'_{ijk(l)} = y_{ijk(l)} - \hat{\mu}$. Since the replicates, rows and columns are orthogonal the estimated effects, ignoring treatment effects, $\hat{\beta}_i, \hat{\rho}_j, \hat{\gamma}_k$, can be computed using the appropriate means of the $y'_{ijk(l)}$, and the unadjusted sum of squares computed as the appropriate sum of squared totals for the $y'_{ijk(l)}$ divided by number of units per total. The observations adjusted for replicates, rows and columns can then be computed by subtracting the estimated effects from $y'_{ijk(l)}$ to give $y''_{ijk(l)}$.

In the case of a Latin square design the treatments are orthogonal to replicates, rows and columns and so the treatment effects, $\hat{\tau}_l$, can be estimated as the treatment means of the adjusted observations, $y''_{ijk(l)}$. The treatment sum of squares is computed as the sum of squared treatment totals of the $y''_{ij(l)}$ divided by the number of times each treatment is replicated. 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 which is not orthogonal, for example a lattice square or an incomplete latin square, the treatment effects adjusted for replicates, rows and columns need to be computed. The adjusted treatment effects are found as the solution to the equations:

$$A\hat{\tau} = \left(R - N_b N_b^{\mathrm{T}}/(rc) - N_r N_r^{\mathrm{T}}/(bc) - N_c N_c^{\mathrm{T}}/(br)\right)\hat{\tau} = q$$

where q is the vector of the treatment totals of the observations adjusted for replicates, rows and columns, $y''_{ijk(l)}$, R is a diagonal matrix with R_{ll} equal to the number of times the lth treatment is replicated, and N_b is the t by b incidence matrix, with $N_{l,i}$ equal to the number of times treatment l occurs in replicate i, with N_r and N_c being similarly defined for rows and columns. The solution to the equations can be written as:

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

where, Ω is a generalized inverse of A. The solution is found from the eigenvalue decomposition of A. The residuals are first calculated by subtracting the estimated adjusted treatment effects from the adjusted observations to give $r'_{ij(l)} = y''_{ij(l)} - \hat{\tau}_l$. However, since only the unadjusted replicate, row and column effects have been removed and they are not orthogonal to treatments, the replicate, row and column means of the $r'_{ij(l)}$ have to be subtracted to give the correct residuals, $r_{ij(l)}$ and residual sum of squares.

Given the sums of squares, the mean squares are computed as the sums of squares divided by the degrees of freedom. The degrees of freedom for the unadjusted replicates, rows and columns are $b - 1$, $r - 1$ and $c - 1$ respectively and for the Latin square 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 Latin square designs, are given by:

$$se(\hat{\tau}_j - \hat{\tau}_{j*}) = \sqrt{2s^2/(bt)}$$

where s^2 is the residual mean square. In the general case the variances of the treatment effects are given by:

$$\text{Var}(\hat{\tau}) = \Omega s^2$$

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

The analysis of a row-column design can be considered as consisting of different strata: the replicate stratum, the rows within replicate and the columns within replicate strata and the units stratum. In the Latin square design all the information on the treatment effects is given at the units stratum. In other designs there may be a loss of information due to the non-orthogonality of treatments and replicates, rows and columns and information on treatments may be available in higher strata. The efficiency of the estimation at the units stratum is given by the (canonical) efficiency factors, these are the nonzero eigenvalues of the matrix, A, 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 information on some treatment comparisons can only be obtained from higher strata.

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

5.1   Compulsory Input Parameters

1:    **nrep – int32 scalar**

b, the number of replicates.

Constraint: **nrep** ≥ 1.

2: **nrow – int32 scalar**

r, the number of rows per replicate.

Constraint: **nrow** ≥ 2.

3: **ncol – int32 scalar**

c, the number of columns per replicate.

Constraint: **ncol** ≥ 2.

4: **y(nrep × nrow × ncol) – double array**

The $n = brc$ observations ordered by columns within rows within replicates. That is $\mathbf{y}(rc(i-1) + r(j-1) + k)$ contains the observation from the kth column of the jth row of the ith replicate for $i = 1, 2, \ldots, b; j = 1, 2 \ldots, r; k = 1, 2, \ldots, c$.

5: **nt – int32 scalar**

the number of treatments. If only replicates, rows and columns are required in the analysis then set **nt** $= 1$.

Constraint: **nt** ≥ 1.

6: **it(∗) – int32 array**

Note: the dimension of the array **it** must be at least **nrep** × **nrow** × **ncol** if **nt** > 1, and at least 1 otherwise.

If **nt** > 1, **it**(i) indicates which of the **nt** treatments unit i received, for $i = 1, 2, \ldots, n$.

If **nt** = 1, **it** is not referenced.

Constraint: $1 \leq \mathbf{it}(i) \leq \mathbf{nt}$ if **nt** ≥ 2, for $i = 1, 2, \ldots, n$.

7: **tol – double scalar**

The tolerance value used to check for zero eigenvalues of the matrix Ω. If **tol** $= 0.0$ a default value of 0.00001 is used.

Constraint: **tol** ≥ 0.0.

8: **irdf – int32 scalar**

An adjustment to the degrees of freedom for the residual and total.

irdf ≥ 1

The degrees of freedom for the total is set to $n - \mathbf{irdf}$ and the residual degrees of freedom adjusted accordingly.

irdf $= 0$

the total degrees of freedom for the total is set to $n - 1$, as usual.

Constraint: **irdf** ≥ 0.

5.2 Optional Input Parameters

None.

5.3 Input Parameters Omitted from the MATLAB Interface

ldtabl, ldc, wk

5.4 Output Parameters

1:    **gmean – double scalar**

   The grand mean, $\hat{\mu}$.

2:    **tmean**(**nt**) **– double array**

   If $\mathbf{nt} \geq 2$, **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 $y_{ijk(l)} - \hat{\tau}_l$. Otherwise **tmean** is not referenced.

3:    **tabl**(**ldtabl,5**) **– double array**

   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 replicates, row 2 for rows, row 3 for columns, row 4 for treatments (if $\mathbf{nt} > 1$), row 5 for residual and row 6 for total. Mean squares are computed for all but the total row, F-statistics and significance are computed for treatments, replicates, rows and columns. Any unfilled cells are set to zero.

4:    **c**(**ldc,nt**) **– double array**

   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 $i = 1, 2, \ldots, t; j = 1, 2, \ldots, t$.

5:    **irep**(**nt**) **– int32 array**

   If $\mathbf{nt} > 1$, the treatment replications, R_{ll}, for $l = 1, 2, \ldots, \mathbf{nt}$. Otherwise **irep** is not referenced.

6:    **rpmean**(**nrep**) **– double array**

   If $\mathbf{nrep} > 1$, **rpmean**(i) contains the mean for the ith replicate, $\hat{\mu} + \hat{\beta}_i$ for $i = 1, 2, \ldots, b$. Otherwise **rpmean** is not referenced.

7:    **rmean**(**nrep** $\times$ **nrow**) **– double array**

   rmean(j) contains the mean for the jth row, $\hat{\mu} + \hat{\rho}_i$ for $j = 1, 2, \ldots, r$.

8:    **cmean**(**nrep** $\times$ **ncol**) **– double array**

   cmean(k) contains the mean for the kth column, $\hat{\mu} + \hat{\gamma}_k$ for $k = 1, 2, \ldots, c$.

9:    **r**(**nrep** $\times$ **nrow** $\times$ **ncol**) **– double array**

   The residuals, r_i for $i = 1, 2, \ldots, n$.

10:   **ef**(**nt**) **– double array**

   If $\mathbf{nt} \geq 2$, the canonical efficiency factors. Otherwise **ef** is not referenced.

11:   **ifail – int32 scalar**

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

6   Error Indicators and Warnings

Note: g04bc may return useful information for one or more of the following detected errors or warnings.

ifail $= 1$

> On entry, **nrep** < 1,
> or   **nrow** < 2,
> or   **ncol** < 2,
> or   **nt** < 1,
> or   **ldtabl** < 6,
> or   **ldc** $<$ **nt**,
> or   **tol** < 0.0,
> or   **irdf** < 0.

ifail $= 2$

> 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 $= 3$

> On entry, the values of **y** are constant.

ifail $= 4$

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

ifail $= 5$

> The treatments are totally confounded with replicates, rows and columns, so the treatment sum of squares and degrees of freedom are zero. The analysis of variance table is not computed, except for replicate, row, column and total sums of squares and degrees of freedom.

ifail $= 6$

> 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 $= 7$

> The design is disconnected, the standard errors may not be valid. The design may have a nested structure.

7   Accuracy

The algorithm used in g04bc, 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. 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, g02cb or g02da. The results from those functions 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 replicates, rows and columns only.

9 Example

```
nrep = int32(1);
nrow = int32(5);
ncol = int32(5);
y = [6.67;
     7.15;
     8.289999999999999;
     8.949999999999999;
     9.619999999999999;
     5.4;
     4.77;
     5.4;
     7.54;
     6.93;
     7.32;
     8.529999999999999;
     8.5;
     9.99;
     9.68;
     4.92;
     5;
     7.29;
     7.85;
     7.08;
     4.88;
     6.16;
     7.83;
     5.38;
     8.51];
nt = int32(5);
it = [int32(5);
      int32(4);
      int32(1);
      int32(3);
      int32(2);
      int32(2);
      int32(5);
      int32(4);
      int32(1);
      int32(3);
      int32(3);
      int32(2);
      int32(5);
      int32(4);
      int32(1);
      int32(1);
      int32(3);
      int32(2);
      int32(5);
      int32(4);
      int32(4);
      int32(1);
      int32(3);
      int32(2);
      int32(5)];
tol = 1e-05;
irdf = int32(0);
[gmean, tmean, table, c, irep, rpmean, rmean, cmean, r, ef, ifail] = ...
    g04bc(nrep, nrow, ncol, y, nt, it, tol, irdf)

gmean =
    7.1856
tmean =
```

```
       7.3180
       7.2440
       7.2060
       6.9000
       7.2600
table =
          0           0           0           0           0
     4.0000     29.4231      7.3558      9.0266      0.0013
     4.0000     22.9950      5.7487      7.0545      0.0037
     4.0000      0.5423      0.1356      0.1664      0.9514
    12.0000      9.7788      0.8149           0           0
    24.0000     62.7392           0           0           0
c =
     0.1304     -0.0326     -0.0326     -0.0326     -0.0326
     0.5709      0.1304     -0.0326     -0.0326     -0.0326
     0.5709      0.5709      0.1304     -0.0326     -0.0326
     0.5709      0.5709      0.5709      0.1304     -0.0326
     0.5709      0.5709      0.5709      0.5709      0.1304
irep =
            5
            5
            5
            5
            5
rpmean =
       0
rmean =
     8.1360
     6.0080
     8.8040
     6.4280
     6.5520
cmean =
     5.8380
     6.3220
     7.4620
     7.9420
     8.3640
r =
    -0.1928
     0.1632
    -0.2548
     0.0372
     0.2472
     0.6812
    -0.4488
    -0.5988
     0.6432
    -0.2768
    -0.1568
     0.5312
    -0.6548
     0.7152
    -0.4348
    -0.2928
    -0.5848
     0.5272
     0.5912
    -0.2408
    -0.0388
     0.3392
     0.9812
    -1.9868
     0.7052
ef =
       1
       1
       1
       1
       1
```

```
ifail =
              0
```