NAG Library Routine Document

G02JAF

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

G02JAF fits a linear mixed effects regression model using restricted maximum likelihood (REML).

2 Specification

```
SUBROUTINE G02JAF (N, NCOL, LDDAT, DAT, LEVELS, YVID, CWID, NFV, FVID,      &
                   FINT, NRV, RVID, NVPR, VPR, RINT, SVID, GAMMA, NFF,      &
                   NRF, DF, REML, LB, B, SE, MAXIT, TOL, WARN, IFAIL)

INTEGER            N, NCOL, LDDAT, LEVELS(NCOL), YVID, CWID, NFV,           &
                   FVID(NFV), FINT, NRV, RVID(NRV), NVPR, VPR(NRV),         &
                   RINT, SVID, NFF, NRF, DF, LB, MAXIT, WARN, IFAIL
REAL (KIND=nag_wp) DAT(LDDAT,NCOL), GAMMA(NVPR+2), REML, B(LB),             &
                   SE(LB), TOL
```

3 Description

G02JAF fits a model of the form:

$$y = X\beta + Z\nu + \epsilon$$

where

y is a vector of n observations on the dependent variable,

X is a known n by p design matrix for the fixed independent variables,

β is a vector of length p of unknown *fixed effects*,

Z is a known n by q design matrix for the random independent variables,

ν is a vector of length q of unknown *random effects*,

and

ϵ is a vector of length n of unknown random errors.

Both ν and ϵ are assumed to have a Gaussian distribution with expectation zero and

$$\mathrm{Var}\begin{bmatrix} \nu \\ \epsilon \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix}$$

where $R = \sigma_R^2 I$, I is the $n \times n$ identity matrix and G is a diagonal matrix. It is assumed that the random variables, Z, can be subdivided into $g \leq q$ groups with each group being identically distributed with expectations zero and variance σ_i^2. The diagonal elements of matrix G therefore take one of the values $\left\{\sigma_i^2 : i = 1, 2, \ldots, g\right\}$, depending on which group the associated random variable belongs to.

The model therefore contains three sets of unknowns, the fixed effects, β, the random effects ν and a vector of $g + 1$ variance components, γ, where $\gamma = \left\{\sigma_1^2, \sigma_2^2, \ldots, \sigma_{g-1}^2, \sigma_g^2, \sigma_R^2\right\}$. Rather than working directly with γ, G02JAF uses an iterative process to estimate $\gamma^* = \left\{\sigma_1^2/\sigma_R^2, \sigma_2^2/\sigma_R^2, \ldots, \sigma_{g-1}^2/\sigma_R^2, \sigma_g^2/\sigma_R^2, 1\right\}$. Due to the iterative nature of the estimation a set of initial values, γ_0, for γ^* is required. G02JAF allows these initial values either to be supplied by you or calculated from the data using the minimum variance quadratic unbiased estimators (MIVQUE0) suggested by Rao (1972).

G02JAF fits the model using a quasi-Newton algorithm to maximize the restricted log-likelihood function:

$$-2l_R = \log(|V|) + (n-p)\log\left(r'V^{-1}r\right) + \log\left|X'V^{-1}X\right| + (n-p)(1+\log(2\pi/(n-p)))$$

where

$$V = ZGZ' + R, \quad r = y - Xb \quad \text{and} \quad b = \left(X'V^{-1}X\right)^{-1}X'V^{-1}y.$$

Once the final estimates for γ^* have been obtained, the value of σ_R^2 is given by:

$$\sigma_R^2 = \left(r'V^{-1}r\right)/(n-p).$$

Case weights, W_c, can be incorporated into the model by replacing $X'X$ and $Z'Z$ with $X'W_cX$ and $Z'W_cZ$ respectively, for a diagonal weight matrix W_c.

The log-likelihood, l_R, is calculated using the sweep algorithm detailed in Wolfinger *et al.* (1994).

4 References

Goodnight J H (1979) A tutorial on the SWEEP operator *The American Statistician* **33(3)** 149–158

Harville D A (1977) Maximum likelihood approaches to variance component estimation and to related problems *JASA* **72** 320–340

Rao C R (1972) Estimation of variance and covariance components in a linear model *J. Am. Stat. Assoc.* **67** 112–115

Stroup W W (1989) Predictable functions and prediction space in the mixed model procedure *Applications of Mixed Models in Agriculture and Related Disciplines* **Southern Cooperative Series Bulletin No. 343** 39–48

Wolfinger R, Tobias R and Sall J (1994) Computing Gaussian likelihoods and their derivatives for general linear mixed models *SIAM Sci. Statist. Comput.* **15** 1294–1310

5 Arguments

1: N – INTEGER *Input*

On entry: n, the number of observations.

Constraint: $\text{N} \geq 1$.

2: NCOL – INTEGER *Input*

On entry: the number of columns in the data matrix, DAT.

Constraint: $\text{NCOL} \geq 1$.

3: LDDAT – INTEGER *Input*

On entry: the first dimension of the array DAT as declared in the (sub)program from which G02JAF is called.

Constraint: $\text{LDDAT} \geq \text{N}$.

4: $\text{DAT}(\text{LDDAT}, \text{NCOL})$ – REAL (KIND=nag_wp) array *Input*

On entry: array containing all of the data. For the ith observation:

$\text{DAT}(i, \text{YVID})$ holds the dependent variable, y;

if $\text{CWID} \neq 0$, $\text{DAT}(i, \text{CWID})$ holds the case weights;

if $\text{SVID} \neq 0$, $\text{DAT}(i, \text{SVID})$ holds the subject variable.

The remaining columns hold the values of the independent variables.

Constraints:

if $\text{CWID} \neq 0$, $\text{DAT}(i, \text{CWID}) \geq 0.0$;
if $\text{LEVELS}(j) \neq 1$, $1 \leq \text{DAT}(i, j) \leq \text{LEVELS}(j)$.

5: LEVELS(NCOL) – INTEGER array *Input*

On entry: LEVELS(i) contains the number of levels associated with the ith variable of the data matrix DAT. If this variable is continuous or binary (i.e., only takes the values zero or one) then LEVELS(i) should be 1; if the variable is discrete then LEVELS(i) is the number of levels associated with it and DAT(j, i) is assumed to take the values 1 to LEVELS(i), for $j = 1, 2, \ldots, \text{N}$.

Constraint: $\text{LEVELS}(i) \geq 1$, for $i = 1, 2, \ldots, \text{NCOL}$.

6: YVID – INTEGER *Input*

On entry: the column of DAT holding the dependent, y, variable.

Constraint: $1 \leq \text{YVID} \leq \text{NCOL}$.

7: CWID – INTEGER *Input*

On entry: the column of DAT holding the case weights.

If $\text{CWID} = 0$, no weights are used.

Constraint: $0 \leq \text{CWID} \leq \text{NCOL}$.

8: NFV – INTEGER *Input*

On entry: the number of independent variables in the model which are to be treated as being fixed.

Constraint: $0 \leq \text{NFV} < \text{NCOL}$.

9: FVID(NFV) – INTEGER array *Input*

On entry: the columns of the data matrix DAT holding the fixed independent variables with FVID(i) holding the column number corresponding to the ith fixed variable.

Constraint: $1 \leq \text{FVID}(i) \leq \text{NCOL}$, for $i = 1, 2, \ldots, \text{NFV}$.

10: FINT – INTEGER *Input*

On entry: flag indicating whether a fixed intercept is included ($\text{FINT} = 1$).

Constraint: $\text{FINT} = 0$ or 1.

11: NRV – INTEGER *Input*

On entry: the number of independent variables in the model which are to be treated as being random.

Constraints:

$0 \leq \text{NRV} < \text{NCOL}$;
$\text{NRV} + \text{RINT} > 0$.

12: RVID(NRV) – INTEGER array *Input*

On entry: the columns of the data matrix DAT holding the random independent variables with RVID(i) holding the column number corresponding to the ith random variable.

Constraint: $1 \leq \text{RVID}(i) \leq \text{NCOL}$, for $i = 1, 2, \ldots, \text{NRV}$.

13: NVPR – INTEGER *Input*

On entry: if $\text{RINT} = 1$ and $\text{SVID} \neq 0$, NVPR is the number of variance components being estimated $- 2$, $(g - 1)$, else $\text{NVPR} = g$.

If $\text{NRV} = 0$, NVPR is not referenced.

Constraint: if $\text{NRV} \neq 0$, $1 \leq \text{NVPR} \leq \text{NRV}$.

14: VPR(NRV) – INTEGER array *Input*

On entry: $\text{VPR}(i)$ holds a flag indicating the variance of the ith random variable. The variance of the ith random variable is σ_j^2, where $j = \text{VPR}(i) + 1$ if $\text{RINT} = 1$ and $\text{SVID} \neq 0$ and $j = \text{VPR}(i)$ otherwise. Random variables with the same value of j are assumed to be taken from the same distribution.

Constraint: $1 \leq \text{VPR}(i) \leq \text{NVPR}$, for $i = 1, 2, \ldots, \text{NRV}$.

15: RINT – INTEGER *Input*

On entry: flag indicating whether a random intercept is included ($\text{RINT} = 1$).

If $\text{SVID} = 0$, RINT is not referenced.

Constraint: $\text{RINT} = 0$ or 1.

16: SVID – INTEGER *Input*

On entry: the column of DAT holding the subject variable.

If $\text{SVID} = 0$, no subject variable is used.

Specifying a subject variable is equivalent to specifying the interaction between that variable and all of the random-effects. Letting the notation $Z_1 \times Z_S$ denote the interaction between variables Z_1 and Z_S, fitting a model with $\text{RINT} = 0$, random-effects $Z_1 + Z_2$ and subject variable Z_S is equivalent to fitting a model with random-effects $Z_1 \times Z_S + Z_2 \times Z_S$ and no subject variable. If $\text{RINT} = 1$ the model is equivalent to fitting $Z_S + Z_1 \times Z_S + Z_2 \times Z_S$ and no subject variable.

Constraint: $0 \leq \text{SVID} \leq \text{NCOL}$.

17: GAMMA(NVPR + 2) – REAL (KIND=nag_wp) array *Input/Output*

On entry: holds the initial values of the variance components, γ_0, with $\text{GAMMA}(i)$ the initial value for σ_i^2/σ_R^2, for $i = 1, 2, \ldots, g$. If $\text{RINT} = 1$ and $\text{SVID} \neq 0$, $g = \text{NVPR} + 1$, else $g = \text{NVPR}$.

If $\text{GAMMA}(1) = -1.0$, the remaining elements of GAMMA are ignored and the initial values for the variance components are estimated from the data using MIVQUE0.

On exit: $\text{GAMMA}(i)$, for $i = 1, 2, \ldots, g$, holds the final estimate of σ_i^2 and $\text{GAMMA}(g + 1)$ holds the final estimate for σ_R^2.

Constraint: $\text{GAMMA}(1) = -1.0$ or $\text{GAMMA}(i) \geq 0.0$, for $i = 1, 2, \ldots, g$.

18: NFF – INTEGER *Output*

On exit: the number of fixed effects estimated (i.e., the number of columns, p, in the design matrix X).

19: NRF – INTEGER *Output*

On exit: the number of random effects estimated (i.e., the number of columns, q, in the design matrix Z).

20: DF – INTEGER *Output*

On exit: the degrees of freedom.

21: REML – REAL (KIND=nag_wp) *Output*

On exit: $-2l_R(\hat{\gamma})$ where l_R is the log of the restricted maximum likelihood calculated at $\hat{\gamma}$, the estimated variance components returned in GAMMA.

22: LB – INTEGER *Input*

On entry: the size of the array B.

Constraint:
$$\mathrm{LB} \geq \mathrm{FINT} + \sum_{i=1}^{\mathrm{NFV}} \max(\mathrm{LEVELS}(\mathrm{FVID}(i)) - 1, 1) + L_S \times \left(\mathrm{RINT} + \sum_{i=1}^{\mathrm{NRV}} \mathrm{LEVELS}(\mathrm{RVID}(i))\right)$$
where $L_S = \mathrm{LEVELS}(\mathrm{SVID})$ if $\mathrm{SVID} \neq 0$ and 1 otherwise.

23: B(LB) – REAL (KIND=nag_wp) array *Output*

On exit: the parameter estimates, (β, ν), with the first NFF elements of B containing the fixed effect parameter estimates, β and the next NRF elements of B containing the random effect parameter estimates, ν.

Fixed effects

If $\mathrm{FINT} = 1$, $\mathrm{B}(1)$ contains the estimate of the fixed intercept. Let L_i denote the number of levels associated with the ith fixed variable, that is $L_i = \mathrm{LEVELS}(\mathrm{FVID}(i))$. Define

if $\mathrm{FINT} = 1$, $F_1 = 2$ else if $\mathrm{FINT} = 0$, $F_1 = 1$;

$F_{i+1} = F_i + \max(L_i - 1, 1)$, $i \geq 1$.

Then for $i = 1, 2, \ldots, \mathrm{NFV}$:

if $L_i > 1$, $\mathrm{B}(F_i + j - 2)$ contains the parameter estimate for the jth level of the ith fixed variable, for $j = 2, 3, \ldots, L_i$;

if $L_i \leq 1$, $\mathrm{B}(F_i)$ contains the parameter estimate for the ith fixed variable.

Random effects

Redefining L_i to denote the number of levels associated with the ith random variable, that is $L_i = \mathrm{LEVELS}(\mathrm{RVID}(i))$. Define

if $\mathrm{RINT} = 1$, $R_1 = 2$ else if $\mathrm{RINT} = 0$, $R_1 = 1$;
$R_{i+1} = R_i + L_i$, $i \geq 1$.

Then for $i = 1, 2, \ldots, \mathrm{NRV}$:

if $\mathrm{SVID} = 0$,

if $L_i > 1$, $\mathrm{B}(\mathrm{NFF} + R_i + j - 1)$ contains the parameter estimate for the jth level of the ith random variable, for $j = 1, 2, \ldots, L_i$;

if $L_i \leq 1$, $\mathrm{B}(\mathrm{NFF} + R_i)$ contains the parameter estimate for the ith random variable;

if $\mathrm{SVID} \neq 0$,

let L_S denote the number of levels associated with the subject variable, that is $L_S = \mathrm{LEVELS}(\mathrm{SVID})$;

if $L_i > 1$, $\mathrm{B}(\mathrm{NFF} + (s-1)L_S + R_i + j - 1)$ contains the parameter estimate for the interaction between the sth level of the subject variable and the jth level of the ith random variable, for $s = 1, 2, \ldots, L_S$ and $j = 1, 2, \ldots, L_i$;

if $L_i \leq 1$, $\mathrm{B}(\mathrm{NFF} + (s-1)L_S + R_i)$ contains the parameter estimate for the interaction between the sth level of the subject variable and the ith random variable, for $s = 1, 2, \ldots, L_S$;

if $\mathrm{RINT} = 1$, $\mathrm{B}(\mathrm{NFF} + 1)$ contains the estimate of the random intercept.

24: SE(LB) – REAL (KIND=nag_wp) array *Output*

On exit: the standard errors of the parameter estimates given in B.

25: MAXIT – INTEGER *Input*

On entry: the maximum number of iterations.

If MAXIT < 0, the default value of 100 is used.

If MAXIT $= 0$, the parameter estimates (β, ν) and corresponding standard errors are calculated based on the value of γ_0 supplied in GAMMA.

26: TOL – REAL (KIND=nag_wp) *Input*

On entry: the tolerance used to assess convergence.

If TOL ≤ 0.0, the default value of $\epsilon^{0.7}$ is used, where ϵ is the ***machine precision***.

27: WARN – INTEGER *Output*

On exit: is set to 1 if a variance component was estimated to be a negative value during the fitting process. Otherwise WARN is set to 0.

If WARN $= 1$, the negative estimate is set to zero and the estimation process allowed to continue.

28: IFAIL – INTEGER *Input/Output*

On entry: IFAIL must be set to 0, -1 or 1. If you are unfamiliar with this argument you should refer to Section 3.4 in How to Use the NAG Library and its Documentation 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 argument, 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, CWID $= \langle value \rangle$ and NCOL $= \langle value \rangle$.
Constraint: $0 \leq$ CWID $\leq$ NCOL and any supplied weights must be ≥ 0.0.

On entry, FINT $= \langle value \rangle$.
Constraint: FINT $= 0$ or 1.

On entry, LB too small: LB $= \langle value \rangle$.

On entry, LDDAT $= \langle value \rangle$ and N $= \langle value \rangle$.
Constraint: LDDAT $\geq$ N.

On entry, N $= \langle value \rangle$.
Constraint: N ≥ 1.

On entry, N < 1 (nonzero weighted observations): N $= \langle value \rangle$.

On entry, NCOL $= \langle value \rangle$.
Constraint: NCOL ≥ 1.

On entry, NFV $= \langle value \rangle$ and NCOL $= \langle value \rangle$.
Constraint: $0 \leq$ NFV $<$ NCOL.

On entry, NRV $= \langle value \rangle$ and NCOL $= \langle value \rangle$.
Constraint: $0 \leq$ NRV $<$ NCOL and NRV $+$ RINT > 0.

On entry, NVPR $= \langle value \rangle$ and NRV $= \langle value \rangle$.
Constraint: $0 \leq$ NVPR $\leq$ NRV and (NRV $\neq 0$ or NVPR ≥ 1).

On entry, RINT $= \langle value \rangle$.
Constraint: RINT $= 0$ or 1.

On entry, SVID $= \langle value \rangle$ and NCOL $= \langle value \rangle$.
Constraint: $0 \leq$ SVID $\leq$ NCOL.

On entry, YVID $= \langle value \rangle$ and NCOL $= \langle value \rangle$.
Constraint: $1 \leq$ YVID $\leq$ NCOL.

IFAIL $= 2$

On entry, GAMMA$(i) < 0.0$, for at least one i.

On entry, invalid data: categorical variable with value greater than that specified in LEVELS.

On entry, LEVELS$(I) < 1$, for at least one I.

On entry, NCOL $= \langle value \rangle$.
Constraint: $1 \leq$ FVID$(i) \leq$ NCOL, for all i.

On entry, NCOL $= \langle value \rangle$.
Constraint: $1 \leq$ RVID$(i) \leq$ NCOL, for all i.

On entry, NVPR $= \langle value \rangle$.
Constraint: $1 \leq$ VPR$(i) \leq$ NVPR, for all i.

IFAIL $= 3$

Degrees of freedom < 1: DF $= \langle value \rangle$.
This is due to the number of parameters exceeding the effective number of observations.

IFAIL $= 4$

Routine failed to converge in MAXIT iterations: MAXIT $= \langle value \rangle$.
See Section 10 for advice.

Routine failed to converge to specified tolerance: TOL $= \langle value \rangle$.
See Section 10 for advice.

IFAIL $= -99$

An unexpected error has been triggered by this routine. Please contact NAG.

See Section 3.9 in How to Use the NAG Library and its Documentation for further information.

IFAIL $= -399$

Your licence key may have expired or may not have been installed correctly.

See Section 3.8 in How to Use the NAG Library and its Documentation for further information.

IFAIL $= -999$

Dynamic memory allocation failed.

See Section 3.7 in How to Use the NAG Library and its Documentation for further information.

7 Accuracy

The accuracy of the results can be adjusted through the use of the TOL argument.

8 Parallelism and Performance

G02JAF is threaded by NAG for parallel execution in multithreaded implementations of the NAG Library.

G02JAF makes calls to BLAS and/or LAPACK routines, which may be threaded within the vendor library used by this implementation. Consult the documentation for the vendor library for further information.

Please consult the X06 Chapter Introduction for information on how to control and interrogate the OpenMP environment used within this routine. Please also consult the Users' Note for your implementation for any additional implementation-specific information.

9 Further Comments

Wherever possible any block structure present in the design matrix Z should be modelled through a subject variable, specified via SVID, rather than being explicitly entered into DAT.

G02JAF uses an iterative process to fit the specified model and for some problems this process may fail to converge (see $\text{IFAIL} = 4$). If the routine fails to converge then the maximum number of iterations (see MAXIT) or tolerance (see TOL) may require increasing; try a different starting estimate in GAMMA. Alternatively, the model can be fit using maximum likelihood (see G02JBF) or using the noniterative MIVQUE0.

To fit the model just using MIVQUE0, the first element of GAMMA should be set to -1.0 and MAXIT should be set to zero.

Although the quasi-Newton algorithm used in G02JAF tends to require more iterations before converging compared to the Newton–Raphson algorithm recommended by Wolfinger *et al.* (1994), it does not require the second derivatives of the likelihood function to be calculated and consequentially takes significantly less time per iteration.

10 Example

The following dataset is taken from Stroup (1989) and arises from a balanced split-plot design with the whole plots arranged in a randomized complete block-design.

In this example the full design matrix for the random independent variable, Z, is given by:

$$Z = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} A & 0 & 0 & 0 \\ 0 & A & 0 & 0 \\ 0 & 0 & A & 0 \\ 0 & 0 & 0 & A \\ A & 0 & 0 & 0 \\ 0 & A & 0 & 0 \\ 0 & 0 & A & 0 \\ 0 & 0 & 0 & A \end{pmatrix}, \tag{1}$$

where

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

The block structure evident in (1) is modelled by specifying a four-level subject variable, taking the values $\{1,1,1,2,2,2,3,3,3,4,4,4,1,1,1,2,2,2,3,3,3,4,4,4\}$. The first column of 1s is added to A by setting $\mathrm{RINT} = 1$. The remaining columns of A are specified by a three level factor, taking the values, $\{1,2,3,1,2,3,1,\ldots\}$.

10.1 Program Text

```
    Program g02jafe

!     G02JAF Example Program Text

!     Mark 26 Release. NAG Copyright 2016.

!     .. Use Statements ..
      Use nag_library, Only: g02jaf, nag_wp
!     .. Implicit None Statement ..
      Implicit None
!     .. Parameters ..
      Integer, Parameter               :: nin = 5, nout = 6
!     .. Local Scalars ..
      Real (Kind=nag_wp)               :: reml, tol
      Integer                          :: cwid, df, fint, i, ifail, j, k, l,   &
```

```
                                        lb, lddat, maxit, n, ncol, nff, nfv, &
                                        nrf, nrv, nvpr, rint, svid, warn,    &
                                        yvid
!     .. Local Arrays ..
      Real (Kind=nag_wp), Allocatable  :: b(:), dat(:,:), gamma(:), se(:)
      Integer, Allocatable             :: fvid(:), levels(:), rvid(:), vpr(:)
!     .. Intrinsic Procedures ..
      Intrinsic                        :: max
!     .. Executable Statements ..
      Write (nout,*) 'G02JAF Example Program Results'
      Write (nout,*)

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

!     Read in the problem size
      Read (nin,*) n, ncol, nfv, nrv, nvpr

      Allocate (levels(ncol),fvid(nfv),rvid(nrv))

!     Read in number of levels for each variable
      Read (nin,*) levels(1:ncol)

!     Read in model information
      Read (nin,*) yvid, fvid(1:nfv), rvid(1:nrv), svid, cwid, fint, rint

!     If no subject specified, then ignore RINT
      If (svid==0) Then
        rint = 0
      End If

!     Calculate LB
      lb = rint
      Do i = 1, nrv
        lb = lb + levels(rvid(i))
      End Do
      If (svid/=0) Then
        lb = lb*levels(svid)
      End If
      lb = lb + fint
      Do i = 1, nfv
        lb = lb + max(levels(fvid(i))-1,1)
      End Do

      lddat = n
      Allocate (vpr(nrv),dat(lddat,ncol),gamma(nvpr+2),b(lb),se(lb))

!     Read in the variance component flag
      Read (nin,*) vpr(1:nrv)

!     Read in the Data matrix
      Read (nin,*)(dat(i,1:ncol),i=1,n)

!     Read in the initial values for GAMMA
      Read (nin,*) gamma(1:(nvpr+rint))

!     Read in the maximum number of iterations
      Read (nin,*) maxit

!     Use default value for tolerance
      tol = 0.0E0_nag_wp

!     Fit the linear mixed effects regression model
      ifail = 0
      Call g02jaf(n,ncol,lddat,dat,levels,yvid,cwid,nfv,fvid,fint,nrv,rvid,    &
        nvpr,vpr,rint,svid,gamma,nff,nrf,df,reml,lb,b,se,maxit,tol,warn,ifail)

!     Display results
      If (warn/=0) Then
        Write (nout,*) 'Warning: At least one variance component was ',       &
          'estimated to be negative and then reset to zero'
```

```
        Write (nout,*)
      End If
      Write (nout,*) 'Fixed effects (Estimate and Standard Deviation)'
      Write (nout,*)
      k = 1
      If (fint==1) Then
        Write (nout,99999) 'Intercept             ', b(k), se(k)
        k = k + 1
      End If
      Do i = 1, nfv
        Do j = 1, levels(fvid(i))
          If (levels(fvid(i))==1 .Or. j/=1) Then
            Write (nout,99995) 'Variable', i, ' Level', j, b(k), se(k)
            k = k + 1
          End If
        End Do
      End Do

      Write (nout,*)
      Write (nout,*) 'Random Effects (Estimate and Standard', ' Deviation)'
      Write (nout,*)
      If (svid==0) Then
        Do i = 1, nrv
          Do j = 1, levels(rvid(i))
            Write (nout,99995) 'Variable', i, ' Level', j, b(k), se(k)
            k = k + 1
          End Do
        End Do
      Else
        Do l = 1, levels(svid)
          If (rint==1) Then
            Write (nout,99998) 'Intercept for Subject Level', l, '         ',  &
              b(k), se(k)
            k = k + 1
          End If
          Do i = 1, nrv
            Do j = 1, levels(rvid(i))
              Write (nout,99997) 'Subject Level', l, ' Variable', i, ' Level', &
                j, b(k), se(k)
              k = k + 1
            End Do
          End Do
        End Do
      End If

      Write (nout,*)
      Write (nout,*) ' Variance Components'
      Write (nout,99996)(i,gamma(i),i=1,nvpr+rint)

      Write (nout,*)
      Write (nout,99994) 'SIGMA^2     = ', gamma(nvpr+rint+1)
      Write (nout,99994) '-2LOG LIKE  = ', reml
      Write (nout,99993) 'DF          = ', df

99999 Format (1X,A,2F10.4)
99998 Format (1X,A,I4,A,2F10.4)
99997 Format (1X,3(A,I4),2F10.4)
99996 Format (1X,I4,F10.4)
99995 Format (1X,2(A,I4),2F10.4)
99994 Format (1X,A,F10.4)
99993 Format (1X,A,I16)
    End Program g02jafe
```

10.2 Program Data

```
G02JAF Example Program Data
24 5 3 1 1
1 4 3 2 3
1 3 4 5 3 2 0 1 1
1
56 1 1 1 1
50 1 2 1 1
39 1 3 1 1
30 2 1 1 1
36 2 2 1 1
33 2 3 1 1
32 3 1 1 1
31 3 2 1 1
15 3 3 1 1
30 4 1 1 1
35 4 2 1 1
17 4 3 1 1
41 1 1 2 1
36 1 2 2 2
35 1 3 2 3
25 2 1 2 1
28 2 2 2 2
30 2 3 2 3
24 3 1 2 1
27 3 2 2 2
19 3 3 2 3
25 4 1 2 1
30 4 2 2 2
18 4 3 2 3
1.0 1.0
-1
```

10.3 Program Results

```
 G02JAF Example Program Results

 Fixed effects (Estimate and Standard Deviation)

 Intercept                37.0000    4.6674
 Variable   1 Level   2    1.0000    3.5173
 Variable   1 Level   3  -11.0000    3.5173
 Variable   2 Level   2   -8.2500    2.1635
 Variable   3 Level   2    0.5000    3.0596
 Variable   3 Level   3    7.7500    3.0596

 Random Effects (Estimate and Standard Deviation)

 Intercept for Subject Level   1            10.7631    4.4865
 Subject Level   1 Variable   1 Level   1    3.7276    3.0331
 Subject Level   1 Variable   1 Level   2   -1.4476    3.0331
 Subject Level   1 Variable   1 Level   3    0.3733    3.0331
 Intercept for Subject Level   2            -0.5269    4.4865
 Subject Level   2 Variable   1 Level   1   -3.7171    3.0331
 Subject Level   2 Variable   1 Level   2   -1.2253    3.0331
 Subject Level   2 Variable   1 Level   3    4.8125    3.0331
 Intercept for Subject Level   3            -5.6450    4.4865
 Subject Level   3 Variable   1 Level   1    0.5903    3.0331
 Subject Level   3 Variable   1 Level   2    0.3987    3.0331
 Subject Level   3 Variable   1 Level   3   -2.3806    3.0331
 Intercept for Subject Level   4            -4.5912    4.4865
 Subject Level   4 Variable   1 Level   1   -0.6009    3.0331
 Subject Level   4 Variable   1 Level   2    2.2742    3.0331
 Subject Level   4 Variable   1 Level   3   -2.8052    3.0331

  Variance Components
    1   62.3958
```

```
   2   15.3819

 SIGMA^2     =     9.3611
 -2LOG LIKE  =   119.7618
 DF          =                16
```