

ESTIMATION OF VARIANCE COMPONENTS AND THEIR SAMPLING VARIANCES

Documentation of Computer Routines

S. R. Searle and R. R. Corbeil
 Biometrics Unit, Cornell University, Ithaca, New York

BU-442-M

Version 10, July 1974

Table of Contents

	<u>Page</u>
[1] Programmers	1
[2] Language	1
[3] Description	1
[4] Models and Methods	2
[5] General Features	3
[6] Sources of Calculation Procedures	3
[7] Optional Output (for Fixed Effects Models)	5
[8] Input	6
[9] Output	7
[10] User Card Deck	7
[11] Description of User Card Deck	8
(a) Control Card	8
(aa) Secondary Control Card	9
(b) User Data Title Card	10
(c) User Factor Title Card	10
[11.1] Data Input when n-Pattern is Known	11
(d) n-Pattern Cards	11
1-way classification	11
2-way nested classification	11
2-way crossed classification	11
3-way nested classification	11
3-way crossed classification	12
(e) Data Cards	12
[11.2] Data Input when n-Pattern is Unknown	13
(f) Data Format Card	13
(g) Data Cards	13
(h) Blank Card	13
(i) Assigned σ^2 Cards	14
(j) Blank Card	14
[12] Example of a User Card Deck	15
[13] Program Flow	16
(a) The MAIN Program	16
(b) Subroutines	16
(c) Overlay Considerations	18
[14] Typographical Corrections to Chapter 11 of LM	20
First Printing	20
Second Printing	21
[15] Availability	22
Acknowledgment	22
References	22
Table 3: Minimum Dimensions for Arrays	23

ESTIMATION OF VARIANCE COMPONENTS AND THEIR SAMPLING VARIANCES

Documentation of Computer Routines

S. R. Searle and R. R. Corbeil

Biometrics Unit, Cornell University, Ithaca, New York

BU-442-M

Version 10, July 1974

[1] Programmers

J. W. Rudan, G. Schaar and C. A. Turturro

[2] Language

ANS FORTRAN IV with IBM 360-370 extensions, suited to the G-level compiler.

[3] Description

This documentation supports a software package developed for computing estimates of variance components and their sampling variances. The package does not require the stricture of equal subclass numbers, i.e., it is designed for handling unequal subclass numbered data (unbalanced data). It can, of course, also handle balanced data but treats them simply as examples of unbalanced data and takes no account of any special features arising from their balanced nature.

Optional output is available for fixed effects models analyses of variance and solution vectors (of the normal equations) for estimating estimable functions.

The estimation procedures used are mostly those given in Chapter 11 of Linear Models, S. R. Searle, Wiley, 1971, which is referred to herein simply as LM. Specific reference to these and other procedures used are given in Section [6].

[4] Models and Methods

Eight different models and seven methods of estimation are provided in the routines. Not all methods of estimation are appropriate to all models. The combinations available are those marked by an asterisk in Table 1.

TABLE 1
 Combinations of Statistical Model and Method of Estimation
 Available in the Computer Routines

CLASS ^{1/} of model	Description	Methods of Estimation						
		Analysis of Variance (Henderson 1)	Henderson 2	Fitting Constants (Henderson 3)	C. R. Rao's MINQUE	An Iterative Method	Maximum Likelihood	Restricted Maximum Likelihood
1.	1-way, random model	*			*		*	*
2.	2-way nested, random model	*					*	*
5.	3-way nested, random model	*					*	*
	<u>2-way crossed</u>							
3.	without interaction, random model	*		*			*	*
4.	with interaction, random model	*		*			*	*
6.	without interaction, mixed model		*	*		*	*	*
7(8).	with interaction, mixed model			*		*	*	*
9.	3-way crossed, random model	*					*	*

^{1/} These CLASS numbers are also used as Control Card input (see Section [11]).

[5] General Features

The program package includes facilities for

- (i) estimating from data both variance components and sampling variances of those estimates,
- (ii) calculating sampling variances of estimated components for pre-assigned values of the population components, and
- (iii) calculating sampling variances of variance component estimators for pre-assigned values of both the population components and the n_{ij} 's — the numbers of observations in the sub-most cells of the data, hereinafter called the n-pattern.

Inasmuch as calculations for variance components estimation include some that are also appropriate to the corresponding fixed effects model, an output option is available for obtaining these calculations for the fixed effects model. This is described in Section [7].

[6] Sources of Calculation Procedures

<u>CLASS</u>		<u>Reference (in LM)</u>
1. 1-way:	ANOVA estimators	Sec. 11.1b
	Covariance matrix of ANOVA estimators	Sec. 11.1c
	Covariance matrix of LSML estimators, and efficiencies	Sec. 11.1d
	MINQUE and MIVQUE estimators	Swallow & Searle [1973]
2. 2-way nested:	ANOVA estimators	Sec. 11.2b
	Covariance matrix of ANOVA estimators	Sec. 11.2c
	Covariance matrix of LSML estimators, and efficiencies	Sec. 11.2d
5. 3-way nested:	ANOVA estimators	Sec. 11.3b
	Covariance matrix of ANOVA estimators	Sec. 11.3c
	Covariance matrix of LSML estimators	Rudan & Searle [1971]

(LSML = Large Sample Maximum Likelihood)

$$\left(\text{Efficiency of ANOVA estimator} = \frac{\text{variance of LSML estimator}}{\text{variance of ANOVA estimator}} \right)$$

Reference (in LM)

3.	2-way crossed, with interaction, random model:	
	ANOVA estimators	Sec. 11.4b
	Covariance matrix of ANOVA estimators	Sec. 11.4c
	Fitting constants estimators	Sec. 11.4e
4.	2-way crossed, no interaction, random model:	
	ANOVA estimators	Sec. 11.6b
	Covariance matrix of ANOVA estimators	Sec. 11.6c
	Fitting constants estimators	Sec. 11.6e
	Covariance matrix of fitting constants estimators	Sec. 11.6f
6.	2-way crossed, no interaction, mixed model:	
	Fitting-constants estimators	Sec. 11.8b ^{1/}
	Covariance matrix of fitting-constants estimators	Sec. 11.6f
	Iterative estimators ^{2/}	Sec. 11.7c ^{1/}
	Henderson Method 2 estimators	Searle [1973]
	Estimators of fixed effects	Sec. 11.8d ^{1/}
7(8).	2-way crossed, with interaction, mixed model:	
	Fitting-constants estimators	Sec. 11.5b ^{1/}
	Iterative estimators ^{2/}	Corbeil & Searle [1973]
	Estimators of fixed effects	" " " "
9.	3-way crossed, with or without any interactions, random model:	
	ANOVA estimators	Sec. 11.9b
For all models:	Maximum likelihood ^{3/} (ML)	Hemmerle & Hartley [1973]
	Restricted maximum likelihood ^{3/} (REML)	Corbeil & Searle [1974]

Note 1: Iterative estimation is limited to either 10 iterations or convergence within 0.001, whichever occurs first.

^{1/} Computing formulae in Searle [1973].

^{2/} Based on Thompson [1969].

^{3/} Large sample covariance matrix also computed.

[7] Optional Output (for Fixed Effects Models)

This output provides basic computations from which calculations for a fixed effects model analysis can be carried out. No attempt is made to provide all the details of a fixed effects model analysis, such as hypothesis testing, F-statistics and so on. That is not the purpose of these routines.

Generally speaking, this optional output consists of 2 parts: (i) degrees of freedom and sums of squares, for use in calculating an analysis of variance table and (ii) cell means and the numbers of observations in each cell, for use in calculating estimates of estimable functions. The output can be used as indicated in the following references. Output can be very extensive with the 2-way crossed classification, especially in the no-interaction case.

This option is not provided for the 3-way crossed classification.

<u>Model</u>	<u>Reference in LM</u>	
	<u>Analysis of variance</u>	<u>Estimates of estimable functions</u>
1-way	Table 6.2, p. 234	p. 235-6
2-way nested	Table 6.7, p. 253	Table 6.8, p. 254
3-way nested	Extension of 2-way nested	
2-way x'd, no interaction	Tables 7.3b and c, p. 275	p. 279
2-way x'd, with interaction	Table 7.8, p. 298	p. 301
3-way crossed	The fixed effects option is not provided	

[8] Input

Six types of input can be used.

- (a) Type 1: observations and bounds on the n-pattern
- (b) Type 2: observations, bounds on the n-pattern, and sets of pre-assigned σ^2 -values
- (c) Type 3: observations and their n-pattern
- (d) Type 4: observations, their n-pattern, and sets of pre-assigned σ^2 -values
- (e) Type 5: an n-pattern and sets of pre-assigned σ^2 -values
- (f) Type 6: n-patterns

Note 2: Input types 1-4 generate estimated components and, where available, sampling variances of those estimates calculated using the estimates. If $\hat{\sigma}^2$ is the vector of estimates and $\text{var}(\hat{\sigma}^2)$ its variance-covariance matrix, then the elements of $\text{var}(\hat{\sigma}^2)$ are quadratic functions of the true variance components, the σ^2 's, the elements of $\underline{\sigma}^2$. Representing these functions by the symbol f we could write

$$\text{var}(\hat{\sigma}^2) = \{f_{ij}(\underline{\sigma}^2)\} \quad \text{for } i, j=1, \dots, t$$

where there are t variance components, including the error variance, in the model. More simply this is the matrix

$$\text{var}(\hat{\sigma}^2) = \underline{F}(\underline{\sigma}^2) \quad . \quad (1)$$

Output for input types 1 and 3 consists of $\hat{\sigma}^2$ and

$$\text{var}(\hat{\sigma}^2)_{\hat{\sigma}^2} = \underline{F}(\hat{\sigma}^2) \quad . \quad (2)$$

Input types 2 and 4 also generate sampling variances

$$\text{var}(\hat{\sigma}^2)_{\sigma_o^2} = \underline{F}(\sigma_o^2) \quad (3)$$

where σ_o^2 represents a set of pre-assigned σ^2 -values.

Input type 5 generates just values of $\text{var}(\hat{\sigma}^2)_{\sigma_o^2}$ of equation (3).

Input type 6 generates just $\text{var}(\hat{\sigma}^2)_{\sigma_o^2}$ for any previously read sets of σ_o^2 -values.

In all cases, up to 150 sets of σ_o^2 -values can be used.

Input types 5 and 6 are available for use in conducting numerical studies of sampling variances of estimators.

[9] Output

The following output is available.

- (a) The n-pattern, whether given as input or derived from the data.
- (b) Sets of pre-assigned values of components, σ_o^2 , given as input.
- (c) Estimates and, where computable, their sampling variances.
- (d) Nested classifications: efficiency.
- (e) Optional output for fixed effects analyses — see Section [7].

[10] User Card Deck

The user card deck is made up as follows.

- (a) Control Card.
- (aa) Secondary Control Card (only for 3-way crossed classification).
- (b) User's data title card.
- (c) User's factor title card.

EITHER

OR

When n-pattern is known (pages 11-12)

When n-pattern is unknown (page 13)

- (d) Cards containing n-pattern.
- (e) Data cards, if required.

- (f) Data format card.
- (g) Data cards.
- (h) One blank card.

- (i) Cards containing pre-assigned σ_o^2 -values, if desired

[Repeated sets of card (a) - (i) as needed.]

- (j) Blank card.

Section [12] shows an example of a user deck.

[11] Description of User Card Deck

The user cards (a) through (j) of the preceding section are now described.

(a) Control Card

The Control Card specifies details of the model being used, the type of input and description of data.

To the left of each value required, sufficient columns are available to (optionally) include the variable name: e.g., column 7 needs 1 digit and columns 1-6 are available for CLASS= . The column to the right of each value must be blank, or contain a comma.

Name Columns

CLASS=i 7 i = 0: Stop

Random models

(Note: These numbers coincide with those used in Sections [4] and [6].)

- 1: 1-way classification
- 2: 2-way nested classification
- 3: 2-way crossed classification, no interaction
- 4: 2-way crossed classification, with interaction
- 5: 3-way nested classification
- 9: 3-way crossed classification, in which case there is a Secondary Control Card (See below).

Mixed models

- 6: 2-way crossed classification, no interaction
- 7: 2-way crossed classification, with interaction
- 8: 2-way crossed classification, with interaction, but without estimating variance components by the iterative method, and without estimating fixed effects at all.

DATA=i 14 i = 1: Input type 1: observations
 2: " " 2: observations and σ^2_{\circ} -values
 3: " " 3: observations and n-pattern
 4: " " 4: observations, n-pattern, and σ^2_{\circ} -values
 5: " " 5: n-pattern and σ^2_{\circ} -values
 6: " " 6: n-pattern

A=i 18-22 i = Number of levels of factor A, range 00001 - 99999

B=i 26-30 i = Maximum number of levels of factor B, range 00000 - 99999 (blank when column 7 contains 1)

C=i 34-38 i = Maximum number of levels of factor C, range 00000 - 99999 (blank unless column 7 contains 5)
See also Notes 3 and 4.

LB=i 43-47 } Blank when DATA (column 14) exceeds 2
When column 14 is 1 or 2

NUB=j 53-57 } i = lower bound and j = upper bound on the codes used for indexing:

- factor A, in a 1-way classification;
- factor B, in a 2-way classification;
- factor C, in a 3-way classification.

(b) User data title card

A single card, containing in columns 1-80, any alphanumeric characters that the user wishes to have as a title to his output. (The output also shows an additional title that identifies the analysis, based on the value given to CLASS in column 7 of the Control Card.)

(c) User factor title card

Columns	1-16	User's name for factor A
"	17-32	User's name for factor B
"	33-48	User's name for factor C

Names need be provided for factors B and C only when used.

[11.1] Data Input when n-Pattern is Known

(d) n-pattern cards

Using n-patterns and pre-assigned σ^2 -values as input provides opportunity for making numerical studies of sampling variances. This is the only purpose for which n-pattern cards are a necessity. The following fixed format in which they must be prepared is therefore not considered a severe limitation on their use.

All n-pattern cards have FORMAT 2013.

Note 5: In all cases a, the number of levels of factor A, has been provided in columns 18-22 of the Control Card. It is therefore not repeated in the n-pattern cards.

1-way classification - factor A (See IM p. 473.)

As many cards as are necessary for containing n_1, n_2, \dots, n_a .

2-way nested classification - factors A, and B within A (See IM p. 475.)

As many cards as are necessary to contain b_1, b_2, \dots, b_a .

(b_i = no. of B levels for i'th level of A.)

a sets of cards: the i'th set has as many cards as are necessary for containing $n_{i1}, n_{i2}, \dots, n_{ib_i}$ for $i = 1, 2, \dots, a$. (When $b_i \leq 20$, the i'th set of cards has 1 card.)

2-way crossed classification - factors A and B. (See IM pp. 480 and 486.)

- in mixed models B is taken as the fixed effects factor.

As many cards as are necessary to contain a b's, i.e., $b_i = b$ for $i = 1, 2, \dots, a$.

a sets of cards: all sets have the same number of cards, as many as are necessary for containing the b values $n_{i1}, n_{i2}, \dots, n_{ib}$ (including 0's) for $i = 1, 2, \dots, a$.

3-way nested classification - factors A, B within A, and C within B. (See IM p.477.)

As many cards as are necessary to contain b_1, b_2, \dots, b_a .

a sets of cards: the i'th set has as many cards as are necessary for containing $c_{i1}, c_{i2}, \dots, c_{ib_i}$ for $i = 1, 2, \dots, a$.

b. sets of cards: the ij'th set has as many cards as are necessary for containing $n_{ij1}, n_{ij2}, \dots, n_{ijc_{ij}}$ for $i = 1, 2, \dots, a$ and $j=1,2,\dots,b_i$.

3-way crossed classification - factors A, B and C. (See LM p. 491.)

As many cards as are necessary to contain a b's, i.e., $b_i = b$ for $i = 1, 2, \dots, a$.

a sets of cards: all sets have the same number of cards, as many as are necessary for containing the b values c, i.e., $c_{ij} = c$ for $j = 1, 2, \dots, b$; and for $i = 1, 2, \dots, a$.

ab sets of cards: ij'th set has as many cards as are necessary for containing the c values $n_{ij1}, n_{ij2}, \dots, n_{ijc}$ (including 0's) for $i = 1, 2, \dots, a$ and $j = 1, 2, \dots, b$; in lexicon order, j within i.

(e) Data cards

Note 6: The numerical studies of sampling variances available from using n-patterns as input require no data (observations). Whenever data for a known n-pattern are available, they can be analyzed either with, or without, using the n-pattern as input. The latter case is the preferred procedure: it yields the n-pattern of the data as analyzed, and it provides variable format options for data input (see page 13). Data input facilities for use in conjunction with n-pattern input are therefore somewhat rigid, i.e., fixed format.

The observations (data) are read under FORMAT(10F8.3) and must be in a sequence that corresponds to the sequence in which the n-pattern has been read. Subscripts to the observations are not needed as input. The first observation of each sub-most cell must begin on a new card, using for each such cell as many cards as are necessary.

Example: If the n_{ij} 's of a 2-way crossed classification are

3	4	2
25	0	7

the data cards would be as follows:

1'st card:	3	observations	for	cell	(1,1)
2'nd	"	4	"	"	(1,2)
3'rd	"	2	"	"	(1,3)
4'th	"	10	"	"	(2,1)
5'th	"	10	"	"	"
6'th	"	5	"	"	"
7'th	"	7	"	"	(2,3)

Note 7: Blank data cards corresponding to empty cells ($n_{ij} = 0$) must not be used.

[11.2] Data Input when n-Pattern is Unknown

(f) Data format card

Columns 1-72 of this card shall contain the FORTRAN FORMAT specification of the data which is to follow. Only that portion of the FORTRAN FORMAT STATEMENT which is in parentheses must appear, inclusive of the parentheses, e.g., (3I5, F8.3). Although this facility provides variable FORMAT, the program does demand that there be only one observation per FORTRAN record, and that there be exactly 4 fields (some of which may be blank). These fields are for I, J, K, and Y_{IJKL} as follows.

(g) Data cards

There shall be one card for each observation, containing I, J, K, and Y_{IJKL} , in that sequence. The last subscript on Y is not needed. The observations for each cell must be sequenced consecutively, and the cells must be in lexicon order, i.e., each index in increasing order and ordered K within J within I.

Example: The format (3I5, F8.3) requires that the data be punched as follows:

	<u>Columns: 1-5</u>	<u>6-10</u>	<u>11-15</u>	<u>16-23 (F8.3)</u>
1-way classification	b	b	I	Y_{IJ}
2-way classification	b	I	J	Y_{IJK}
3-way classification	I	J	K	Y_{IJKL}

b = blank

The codes used for each index, I, J and K, need not be consecutive numbers. Nor need they start at unity. All that is necessary is that the codes be sequenced, in lexicon order, and that upper bounds on I, J and K be given in columns 18-22, 26-30 and 34-38 respectively of the Control Card, as indicated in Note 3. If the codes of any index are not consecutive and are sparsely spread over a large range, the program may call for such large storage requirements as to make it more economical for the user to first recode each such index.

(h) Blank card

When the n-pattern is not being read in, the data must be followed by one blank card, or as many blank cards as there are FORTRAN records required by the FORMAT of (f) above.

(i) Assigned σ^2 cards (σ^2 -values)

Each card contains one set of pre-assigned σ^2 -values, some of which may be zero, depending on the model being used, as shown in the following table. The FORMAT for each card is 4F5.1.

FORMAT 4F5.1

Model	CLASS (Column 7 of Control Card)	Columns			
		1-5	6-10	11-15	16-20
1-way	1	0	0	σ^2_α	σ^2_e
2-way nested	2	0	σ^2_α	σ^2_β	σ^2_e
2-way crossed, with interaction	4,7,8	σ^2_α	σ^2_β ^{1/}	σ^2_γ	σ^2_e
2-way crossed, no interaction	3,6	σ^2_α	σ^2_β ^{1/}	0	σ^2_e
3-way nested	5	σ^2_α	σ^2_β	σ^2_γ	σ^2_e
3-way crossed	9		not available		

^{1/} Zero in mixed models without interaction (CLASS=6), and not used in mixed models with interaction (CLASS=7 or 8), for which sampling variances are not available.

For whichever model is being used there will be as many σ^2 -cards as is indicated in columns 54-56 of the Control Card.

Repetitions of (a) - (i): Successive uses of the program can be made by having cards (a) - (i) for each use follow one another in succession.

(j) Blank card

This finishes the job, which may consist of several uses.

[12] Example of a User Card Deck

Letters in parentheses refer to paragraphs in Sections [10] and [11].

Output

- | | | |
|---|---|---|
| (a), (b), and (c) Control and title cards | } | n-pattern, $\hat{\sigma}^2$'s and their variances using $\hat{\sigma}^2$ and σ_o^2 |
| (f) Data format card | | |
| (g) Data, type 2 of paragraph 8 | | |
| (h) Blank card | | |
| (i) σ_o^2 -values | | |
| (a), (b), and (c) Control and title cards | } | v($\hat{\sigma}^2$) for new n-pattern, using preceding σ_o^2 's. (May be new model, via Control Card, but σ_o^2 's must conform to that model.) |
| (d) n-pattern | | |
| (a), (b), and (c) Control and title cards | } | Same as preceding |
| (d) n-pattern | | |
| (a), (b), and (c) Control and title cards | } | v($\hat{\sigma}^2$) for new n-pattern and σ_{oo}^2 . (May be new model, via Control Card, but σ_{oo}^2 's must conform to that model.) |
| (d) n-pattern | | |
| (i) σ^2 -values, σ_{oo}^2 | | |
| (j) Blank card | | Finish. |

[13] Program Flow

(a) The MAIN Program

The mainline (calling) program, which is in 3 parts, is compiled each time a job is processed because it designates dimensions for data arrays which depend on job size. One scratch file is needed for unformatted data (DATA types 1 or 2) which are referenced in a sequential manner, and for CLASS = 3 or 4 when $B > A$ (see UNFORM and NTRAN described in b below). Another scratch file is used for storing the vector of observations when CLASS = 6 or $ML > 0$.

MAIN: This mainline routine contains all arrays which are dimensioned by the user. It also reads the Control Card(s) identifying the segment of the program package needed for the analysis. Through calling SUBML, MAIN calls the subroutines UNFORM, INFORM, FIXER, STARTN, STARTC, STARTM, INPUT and CWAY; and through SUBM2 it calls PRELNK, LINKUP, MLES and TIML.

(b) Subroutines

UNFORM: Called from MAIN when DATA = 1 or 2. It requires scratch files with data sets reference of 1 and 2. Scratch file 1 should be of a size to contain the vector of all observations. Scratch file 2 should be of a size to contain $(NUB - LB + 1)$ records (columns 43-57 of Control Card).

INFORM: Same as UNFORM but only when CLASS = 9 and DATA = 1.

FIXER: Called from MAIN for use with UNFORM and INFORM. It sets in agreement the dimensions of the n-pattern derived in UNFORM and INFORM with those dimensions required by all other routines.

STARTN: Called by MAIN for nested (including 1-way) classifications, i.e., for CLASS = 1, 2 or 5. Reads input data from a "card reader" (data set reference 5) if not previously read by UNFORM. Calls NESTED, and uses scratch file 1 (see STARTM).

NESTED: Called by STARTN. Calculates variance components for nested (including 1-way) classifications. Calls CDI, MINQ and OUTDAN.

CDI: Called by NESTED, CROSSED, MIXED, MLES and TIML. This is a matrix inversion routine, using Choleski decomposition, to efficiently invert large order, positive-definite, symmetric matrices.

MINQ: Called by NESTED only for CLASS = 1. Calculates MIVQUE and MINQUE estimates and their sampling variances.

OUTDAN: Called by NESTED for printing (data set reference 6) output.

STARTC: Called by MAIN when CLASS = 3 or 4. Reads input data from data set 5 if not previously read by UNFORM. Calls CROSSD and NTRAN, and uses scratch file 1 (see STARTM).

NTRAN: Called by STARTC when B of control card cols. 26-30 is greater than A of cols. 18-22. Effectively reduces the size of certain matrices that need inverting. Requires a scratch file with data set reference 2 and should be of a size to contain AXB records.

CROSSD: Called by STARTC and calculates components of variance for the 2-way crossed classifications. Calls CDI (see above), VARMAT, INV2, MATMUL and OUTDAC.

VARMAT: Called by CROSSD and MIXED. Calculates sampling covariance matrix for analysis of variance estimators.

INV2: Called by CROSSD, MIXED, CWAY, MLES and TIML to solve systems of linear equations and invert non-symmetric matrices.

MATMUL: Called by CROSSD and MIXED. Matrix multiplication without destroying original matrices.

OUTDAC: Called by CROSSD for printing (data set reference 6) output.

STARTM: Called by MAIN when CLASS = 6, 7 or 8. It is similar to STARTN and STARTC but for mixed models where B of Control Card columns 26-30 is the number of fixed effects. It calls MIXED. The scratch file (data set reference 1) should be of a size to accommodate all the observations.

MIXED: Called by STARTM and calculates components of variance of 2-way crossed classifications as though the model was first random and then mixed. Calls CDI, VARMAT, INV2, MATMUL (see above) and OUTDAM.

OUTDAM: Called by MIXED for printing (data set reference 6) output.

- INPUT: Called by MAIN but only for CLASS = 9 and DATA = 3.
- CWAY: Called by MAIN and calculates variance components for 3-way crossed classifications, with optional interactions as per Secondary Control Card; and gives all output. Calls INV2 (see above) and MOVE.
- MOVE: Called by CWAY. Compresses work space in conformity with Secondary Control Card.
- PRELNK: Called by MAIN and calculates constants and dimensions for MLES and TIML.
- LINKUP: Called by MAIN and prepares input for MLES and TIML, reading vector of all observations from data set reference 1.
- MLES: Called by MAIN and calculates maximum likelihood estimates of the variance components, the fixed effects, and their sampling variances.
- TIML: Called by MAIN and calculates restricted maximum likelihood estimates of the variance components, the fixed effects, and their sampling variances.

(c) Overlay Considerations

Coding for the entire program, including arrays used as work spaces, requires approximately 250K of core. This excludes data storage in the arrays listed in Table 3. (See pages 23—25.)

When an overlay feature is available in a computing facility, it can reduce the program requirement from 250K bytes to approximately 100K bytes. A convenient partitioning is shown in Table 2.

TABLE 2
Suggested Overlay Segments

Approximate partition size (bytes)	Segment number	Assigned routines
47K	1 (root phase)	MAIN, UNIFORM, FIXER, CDI, INV2, VARMAT, MATMUL
44K	2	STARTN, NESTED, OUTDAN, MINQ
45K	3	STARTC, NTRAN, CROSSD, OUTDAC
48K	4	STARTM, MIXED, OUTDAM
25K	5	INFORM, INPUT, CWAY
32K	6	PRELNK, LINKUP, MLE, TIML

The maximum data collection which can then be processed is determined by the memory size of the computing facility less 100K bytes for the program coding. Table 3 provides exact specifications for each of the 9 CLASSES of model, from which maximum sized data sets can be calculated.

Storage of individual observations in core is avoided. For all models, each data cell consists of the sum of an unlimited number of replicated observations. The number of cells depends on the model and on A, B, C of the Control Card (columns 18-38).

[14] Typographical Corrections to Chapter 11 of IM

First Printing

Page 474, line 2:

$$T_A \text{ should be } T_A = \sum_{i=1}^a y_{i.}^2 / n_i .$$

Page 474, last line:

$$+t(T_0 - T_A) \text{ should be } -t(T_0 - T_A) .$$

Page 480, Table 11.1:

$$k_{15} \text{ should be } k_{15} = \sum_i (\sum_j n_{ij} n_{.j})^2 / n_i .$$

$$k_{28} \text{ should be } k_{28} = \sum_i \sum_j n_{ij}^4 / n_i \cdot n_{.j} .$$

Page 481, $\hat{\sigma}_Y^2$ before section c:

$$(N - k_1) \delta_B \text{ should be } (N - k'_1) \delta_B .$$

$$\text{last term} + k_{23}) \text{ should be } + k'_{23}) .$$

Page 481, between lines 4-5 up, add:

$$\text{and } \text{cov}(\hat{\sigma}^2, \hat{\sigma}_e^2) = -P^{-1} f v(\hat{\sigma}_e^2) \text{ from (46) on page 434.}$$

Page 482, Title to first row:

$$2\sigma_\alpha \text{ should be } 2\sigma_\alpha^4$$

Page 483, line 9 up:

$$h_7 = N - h_4 = h_8 \text{ should be } h_7 = N - k_4 = h_8$$

Page 488, lines 5-7: The sentence containing $\text{var}(\hat{\sigma}^2)$ should read as follows:

Therefore

$$\text{var}(\hat{\sigma}^2) = Q^{-1} H \text{ var}(t) H' Q^{-1} + 2q_3 q_3' \sigma_e^4 (N - s)$$

where q_3 is the 3rd column of Q^{-1} .

Page 488, line 3 up:

$$-(a + b) \hat{\sigma}_e^2 \text{ should be } -(a + b - 2) \hat{\sigma}_e^2 .$$

Second Printing

- 483/6 up: after following add based on Tables 10.1 and 10.2
- 483/5 up: on right margin add [10.2]
- 483/4 up: on right margin add [10.1]
- 483/3 up: on right margin add [10.1 and 10.2]
- 484/5 up on left side of table 11.3: $j \neq j'$ should be j, j'
- 484/3 up of left side of table 11.3: $t_B = \underset{\sim}{r}' \underset{\sim}{C}^{-1} \underset{\sim}{r}$ should be $t_B = \underset{\sim}{r}' \underset{\sim}{C}^{-1} \underset{\sim}{r} = R(\beta | \mu, \alpha)$
- 486/last equation of section b: $(N - k_4')$ should be $(N - k_4)$
- 488/6: $q_3 q_3'$ should be bold face
- 488/7: q_3 should be bold face
- 490/8: $(\underset{\sim}{X} \underset{\sim}{T} \underset{\sim}{X})^{-1}$ should be $(\underset{\sim}{X}' \underset{\sim}{T} \underset{\sim}{X})^{-1}$
- 491/5: $\sigma_e^2 / \sigma_\mu^2$ should be $\sigma_e^2 / \sigma_\alpha^2$

The list of corrections to the complete book is available from the author.

[15] Availability

The program and additional copies of this documentation (paper BU-442-M) are available from the Biometrics Unit, Warren Hall, Cornell University, Ithaca, N. Y. The charge is \$15. The program is on tape in card images, containing FORTRAN IV source statements, the tape being 9-track, 800 BPI, record length = 80, 10 records per block, with no label. The tape will be accompanied by a summary description (paper BU-492-M) that includes examples of the output.

Acknowledgment

This project was supported by Grant No. GJ-31746 of the National Science Foundation.

References

- Corbeil, R. R. and S. R. Searle (1973). An iterative procedure for estimating variance components in the 2-way crossed classification mixed model, with interaction, using unbalanced data. Paper BU-460-M of the Biometrics Unit, Cornell University.
- Corbeil, R. R. and S. R. Searle (1974). Restricted maximum likelihood estimators of variance components in the mixed models. Paper BU-496-M of the Biometrics Unit, Cornell University.
- Hemmerle, W. J. and H. O. Hartley (1973). Computing maximum likelihood estimates for the mixed A.O.V. model using the W-transformation. Technometrics, 15, 819-831.
- Rudan, J. W. and S. R. Searle (1971). Large sample variances of maximum likelihood estimators of variance components in the 3-way nested classification, random model, with unbalanced data. Biometrics, 27, 1087-1091.
- Searle, S. R. (1971). Linear Models, Wiley, New York.
- Searle, S. R. (1973). Computing procedures for estimating variance components from unbalanced data in the 2-way crossed classification, no interaction, mixed model. Paper BU-450-M of the Biometrics Unit, Cornell University.
- Swallow, W. H. and S. R. Searle (1973). Minimum norm and minimum variance quadratic unbiased estimation of variance components from unbalanced data in the 1-way classification. Paper BU-447-M of the Biometrics Unit, Cornell University.
- Thompson, R. (1969). Iterative estimation of variance components for non-orthogonal data. Biometrics, 25, 767-773.

TABLE 3 (Continued)

Use	Variable name	{1}	{2}	{3,4}	{6}	{7,8}	{5}	{9}
4. Mixed model with interaction	SUMIJ(a)	a=1	a=1	a=1	a=1	a=AXB ^{2/}	a=1	a=1
	Q1(a)	a=1	a=1	a=1	a=1	a=S	a=1	a=1
	QP(a)	a=1	a=1	a=1	a=1	a=S	a=1	a=1
	QPSU(a)	a=1	a=1	a=1	a=1	a=S	a=1	a=1
	D(a)	a=1	a=1	a=1	a=1	a=BX ^{2/}	a=1	a=1
	DP(a)	a=1	a=1	a=1	a=1	a=BX ^{2/}	a=1	a=1
	DS(a)	a=1	a=1	a=1	a=1	a=BX ^{2/}	a=1	a=1
	DSD(a)	a=1	a=1	a=1	a=1	a=SXS ^{2/}	a=1	a=1
	WTW(a)	a=1	a=1	a=1	a=1	a=SXS ^{2/}	a=1	a=1
5. n-pattern	NN(a)	a=A	a=AXB	a=AXB	a=AXB	a=AXB	a=AXBXC	a=AXBXC
	NTJ(a)	a=A	a=A	a=A	a=A	a=A	a=AXB	a=AXB
	NTK(a)	a=1	a=B	a=B	a=B	a=B	a=AXC	a=AXC
	MC(a)	a=1	a=1	a=1	a=1	a=1	a=1	a=C
	NTI(a)	a=1	a=1	a=1	a=1	a=1	a=A	a=A
	JB(a)	a=1	a=1	a=A	a=1	a=1	a=A	a=B
	KC(a)	a=1	a=B	a=B	a=B	a=B	a=AXC	a=A
	IRINV(a)	a=2	a=3	a=B	a=B	a=S	a=4	a=B
	NBC(a)	a=1	a=1	a=1	a=1	a=1	a=1	a=BX ^{2/} C
	NC(a)	a=1	a=1	a=1	a=1	a=1	a=1	a=C

III SUBM2

COMMON/AREA/

For all classifications.^{5/} All are one-dimensional arrays.

E(c)
 DR(c)
 ALPHA(k)
 RR(m+k+1)²
 W(m+k+1)²
 WM(m²) or (k+1)² whichever is larger
 WM²(c×m) or c×(k+1) " " "

Footnotes overleaf.

Footnotes to TABLE 3

- 1/ S is the number of filled cells in crossed classifications ($S \leq A \times B$).
- 2/ $a=1$, for CLASS = 8.
- 3/ The fitting constants procedure sets $\beta = \min(A, B)$.
- 4/ $a=D=1$ except when DATA=1 or 2 of Control Card column 14 then $D = \text{NUB} - \text{LB} + 1$ of columns 43-57.
- 5/ c = number of variance components to be estimated, other than the error variance; k = number of fixed effects (usually B or 1); m = total number of levels in all random factors. When $ML = 0$ (Control Card, column 75) all arrays under III, SUBM2, are assigned dimension 1.
- 6/ COMMON/AREA/ is specified in the 2 subroutines of the mainline in an effort to save storage space. Therefore, SIZE(a) should have sufficiently large dimension to accommodate all arrays specified under SUBM1 or SUBM2, whichever is larger. Unless $ML=0$, SUBM2 will usually have the greater data storage requirements, which can be calculated as the sum of the seven arrays under III multiplied by 8 (8 bytes per stored datum), namely

$$8[3m^2 + m(c+4k+4) + 2c + k + 2(k+1)^2] \quad \text{if} \quad m > k + 1$$

or

$$8[2m^2 + 4m(k+1) + c(2+k+1) + 2(k+1)^2] \quad \text{if} \quad m \leq k + 1 .$$

Likewise, the sum of the array dimensions in SUBM1 can be calculated as the sum of groups 1-4 multiplied by 8 plus the sum of group 5 multiplied by 2. That is, for

$$\text{CLASS}=1: \quad 44A + 16D + 258.$$

$$2: \quad 10AB + 74A + 20B + 16D + 208.$$

$$3,4: \quad 24\beta^2 + 16\beta + 2AB + 12A + 38B + 16D + 144 + 8 \max(A, B).$$

$$5: \quad 10ABC + 74AB + 12AC + 12A + 8C + 16D + 206.$$

$$6: \quad 24B^2 + 2AB + 34A + 78B + 16D + 106.$$

$$7: \quad 16S^2 + 24B^2 + 24BS + 10AB + 34S + 18A + 44B + 16D + 74.$$

$$8: \quad 24B^2 + 2AB + 10S + 18A + 44B + 16D + 146.$$

$$9: \quad 2ABC + 10AB + 10AC + 10BC + 12A + 12B + 12C + 16D + 960 \\ + 8 \max(n_{ijk}).$$