Annotated Computer Output for Analyses of Unbalanced Data:
SAS GLM

BU-641-M by February, 1978

S. R. Searle and H. V. Henderson
Biometrics Unit, Cornell University, Ithaca, New York

Abstract

The material presented here is part of a continuing project designed for producing annotated copies of statistical computing package output, for analysis of variance programs used on unbalanced data. The entire project includes the running of seven data sets on a variety of programs. This report is a preliminary report of running three of those data sets on the GLM routine in SAS. It includes description of the data sets, and copies of annotated output.

Comments

The data sets are entirely hypothetical and are designed solely as vehicles for ascertaining features of computer output. The annotated output should be read in the same sequence as the data sets are numbered. Many comments made on the output of data set 2, for example, are not repeated on data sets 3 and 4 even though applicable. References such as LM 283 are to "Linear Models", S. R. Searle, Wiley and Sons, 1971, – in this case to page 283.

Paper No. BU-641-M in the Biometrics Unit Mimeo Series.

Data Set (2): Unbalanced Data, 0 or 1 observation

2-way crossed classification without interaction,
4 rows and 3 columns [LM, pp. 262, 272]

Row	Column A	Column B	Column C	Total	No. of Observations	Mean
X	18	12	24	54	(3)	18
Y	-	-	9	9	(1)	9
Z	3	-	15	18	(2)	9
W	6	3	18	27	(3)	9
Total	27	15	66	108		
No. of observations	(3)	(2)	(4)		(9)	
Mean	9	$7\frac{1}{2}$	$16\frac{1}{2}$			12

Sums of Squares

$R(\mu)$	= 1296
$R(\mu,\alpha)$	= 1458
$R(\mu,\beta)$	= $1444\frac{1}{2}$
$R(\mu,\alpha,\beta)$	= 1716
$y'y$	= 1728

Analyses of Variance

a. Rows before columns: Term	d.f.	S.S.	b. Columns before rows: Term	d.f.	S.S.
$R(\mu)$	1	1296	$R(\mu)$	1	1296
$R(\alpha\|\mu)$	3	162	$R(\beta\|\mu)$	2	$148\frac{1}{2}$
$R(\beta\|\mu,\alpha)$	2	258	$R(\alpha\|\mu,\beta)$	3	$271\frac{1}{2}$
SSE	3	12	SSE	3	12
SST	9	1728	SST	9	1728

Parameter vector: $\underset{\sim}{b}' = [\mu \quad \alpha_1 \quad \alpha_2 \quad \alpha_3 \quad \alpha_4 \quad \beta_1 \quad \beta_2 \quad \beta_3]$

Solution vector: $\underset{\sim}{b}^{o\prime} = [0 \quad 26 \quad 9 \quad 14 \quad 17 \quad -10 \quad -14 \quad 0]$

Data Set (3): Unbalanced Data, all cells filled

2-way crossed classification, 2 rows and 3 columns
[BU-608-M, pages 4-7]

				Totals
	7 9 16(2)8*	6 6(1)6	2 2(1)2	24(4)6
	8 8(1)8	4 8 12(2)6	12 12(1)12	32(4)8
Totals:	24(3)8	18(3)6	14(2)7	56(8)7

* In each triplet of numbers, the first is a total, the second (in parentheses) is the number of observations, and the third is the mean.

Analyses of Variance

Sums of Squares		a. Rows before columns			b. Columns before rows		
		Term	d.f.	S.S.	Term	d.f.	S.S.
$R(\mu)$	$= 392$	$R(\mu)$	1	392	$R(\mu)$	1	392
$R(\mu,\alpha)$	$= 400$	$R(\alpha\|\mu)$	1	8	$R(\beta\|\mu)$	2	6
$R(\mu,\beta)$	$= 398$	$R(\beta\|\mu,\alpha)$	2	$11\frac{7}{11}$	$R(\alpha\|\mu,\beta)$	1	$13\frac{7}{11}$
$R(\mu,\alpha,\beta)$	$= 411\frac{7}{11}$	$R(\gamma\|\mu,\alpha,\beta)$	2	$36\frac{4}{11}$	$R(\gamma\|\mu,\alpha,\beta)$	2	$36\frac{4}{11}$
$R(\mu,\alpha,\beta,\gamma)$	$= 448$	SSE	2	10	SSE	2	10
$\underset{\sim}{y}'\underset{\sim}{y}$	$= 458$	SST	8	458	SST	8	458

Weighted squares of means: $SSA_w = 20$

$SSB_w = 5\frac{1}{3}$

Parameter vector: $\underset{\sim}{b}' = [\mu \quad \alpha_1 \quad \alpha_2 \quad \beta_1 \quad \beta_2 \quad \beta_3 \quad \gamma_{11} \quad \gamma_{12} \quad \gamma_{13} \quad \gamma_{21} \quad \gamma_{22} \quad \gamma_{23}]$

Solution vector: $\underset{\sim}{b}^{o\prime} = [0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 8 \quad 6 \quad 2 \quad 8 \quad 6 \quad 12]$

Data Set (4): Unbalanced Data, empty cells - a simple case

2-way crossed classification, 2 rows and 3 columns

				Totals
	2 4 6	4 6	5	
	12(3)4	10(2)5	5(1)5	27(6)4½
	12 8	11 7	-	
	20(2)10	18(2)9	-	38(4)9½
Totals:	32(5)6.4	28(4)7	5(1)5	65(10)6½

Analyses of Variance

Sums of Squares		a. Rows before columns: Term	d.f.	S.S.	b. Columns before rows: Term	d.f.	S.S.
$R(\mu)$	= $422\frac{1}{2}$	$R(\mu)$	1	422.5	$R(\mu)$	1	422.5
$R(\mu,\alpha)$	= $482\frac{1}{2}$	$R(\alpha\|\mu)$	1	60	$R(\beta\|\mu)$	2	3.3
$R(\mu,\beta)$	= 425.8	$R(\beta\|\mu,\alpha)$	2	.318	$R(\alpha\|\mu,\beta)$	1	57.018
$R(\mu,\alpha,\beta)$	= 482.818	$R(\gamma\|\mu,\alpha,\beta)$	1	2.182	$R(\gamma\|\mu,\alpha,\beta)$	1	2.182
$R(\mu,\alpha,\beta,\gamma)$	= 485	SSE	5	26	SSE	5	26
$\underset{\sim}{y}'\underset{\sim}{y}$	= 511	SST	10	511	SST	10	511

Parameter vector: $\underset{\sim}{b}' = [\mu \quad \alpha_1 \quad \alpha_2 \quad \beta_1 \quad \beta_2 \quad \beta_3 \quad \gamma_{11} \quad \gamma_{12} \quad \gamma_{13} \quad \gamma_{21} \quad \gamma_{22}]$

Solution vector: $\underset{\sim}{b}^{o'} = [0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 4 \quad 5 \quad 5 \quad 10 \quad 9\ ]$

DATA SET 2
UNBALANCED DATA, NO INTERACTION, N(I,J) = 0 OR 1
LINEAR MODELS BY S R SEARLE PAGE 262

SAS GLM Data Set 2

GENERAL LINEAR MODELS PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
A	4	1X 2Y 3Z 4W
B	3	A B C

NUMBER OF OBSERVATIONS IN DATA SET = 9

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

GENERAL FORM OF ESTIMABLE FUNCTIONS

EFFECT		COEFFICIENTS
INTERCEPT		L1
A	1X	L2
	2Y	L3
	3Z	L4
	4W	L1-L2-L3-L4
B	A	L6
	B	L7
	C	L1-L6-L7

(INTERCEPT) Unfortunate word: mu is preferable.

The coefficients in this output are based on the general result (LM 186) that

$$\tilde{w}'\tilde{H}\tilde{b} \text{ is estimable for } \underline{\text{any}}\ \tilde{w}' ,$$

where

$$\tilde{H} = \tilde{G}\tilde{X}'\tilde{X}', \quad \text{for} \quad \tilde{X}'\tilde{X}\tilde{G}\tilde{X}'\tilde{X} = \tilde{X}'\tilde{X} .$$

For this example, $\tilde{G}$ at bottom of LM 269 and $\tilde{X}'\tilde{X}$ on LM 264 gives

$$\tilde{H} = \begin{bmatrix} . & . & . & . & . & . & . & . \\ 1 & 1 & . & . & . & . & . & 1 \\ 1 & . & 1 & . & . & . & . & 1 \\ 1 & . & . & 1 & . & . & . & 1 \\ 1 & . & . & . & 1 & . & . & 1 \\ . & . & . & . & . & 1 & . & -1 \\ . & . & . & . & . & . & 1 & -1 \\ . & . & . & . & . & . & . & . \end{bmatrix} \quad \text{and} \quad (\tilde{w}'\tilde{H})' = \begin{bmatrix} w_2 + w_3 + w_4 + w_5 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \\ w_6 \\ w_7 \\ w_2 + w_3 + w_4 + w_5 - w_6 - w_7 \end{bmatrix}$$

Putting

$$\begin{aligned} w_2 + w_3 + w_4 + w_5 &= L_1 & w_6 &= L_6 \\ w_2 &= L_2 & w_7 &= L_7 \\ w_3 &= L_3 & w_2 + \cdots + w_5 - w_6 - w_7 &= L_1 - L_6 - L_7 \\ w_4 &= L_4 \\ w_5 &= L_1 - L_2 - L_3 - L_4 \end{aligned}$$

gives the SAS COEFFICIENTS for the general form of estimable function.

The estimable functions described as Types I-III are functions that form the basis of. and can be used for, calculating hypotheses that are tested by F-statistics that have pre-ordained sums of squares as the numerator sum of squares. That is, for each sum of squares that might get used in a numerator in an F-statistic. the output gives $f = \tilde{\ell}'\tilde{b}$ where the composite hypothesis being tested can then be expressed as

$$H: \ f_i = 0 \quad \text{for } i = 1, \cdots, r$$

where r is degrees of freedom, and for r linearly independent estimable functions that are available from $f = \tilde{\ell}'\tilde{b}$. The purpose of Type IV functions differs from that of Types I-III. It is explained on output page 56 of Data Set 4.

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

TYPE I ESTIMABLE FUNCTIONS FOR: A

EFFECT		COEFFICIENTS
INTERCEPT		0
A	1X	L2
	2Y	L3
	3Z	L4
	4W	-L2-L3-L4
B	A	-0.3333*L3+0.1667*L4
	B	-0.3333*L3-0.3333*L4
	C	0.6667*L3+0.1667*L4

For TYPE I, the sums of squares explained are those corresponding to the sequential fitting of the factors as indicated in the data input. For example, if factors A, B, C are entered in that order, the sums of squares explained are $R(\alpha|\mu)$, $R(\beta|\mu,\alpha)$ and $R(\gamma|\mu,\alpha,\beta)$.

Example: This output is for explaining $R(\alpha|\mu)$ in the sequence A, B. From LM 283, $R(\alpha|\mu)$ is used for testing

$$H: \alpha_i + \frac{1}{n_{i.}} \sum_{j=1}^{b} n_{ij}\beta_j \text{ all equal.}$$

n_{ij}'s			
1	1	1	3
-	-	1	1
1	-	1	2
1	1	1	3
3	2	4	9

Hence H is

$$H: \begin{Bmatrix} \alpha_1 + \frac{1}{3}(\beta_1 + \beta_2 + \beta_3) \\ \alpha_2 + \frac{1}{1}(\qquad\qquad \beta_3) \\ \alpha_3 + \frac{1}{2}(\beta_1 \qquad + \beta_3) \\ \alpha_4 + \frac{1}{3}(\beta_1 + \beta_2 + \beta_3) \end{Bmatrix} \text{ equal.}$$

From these estimable functions we form a general contrast of the α's plus a "mess" of other parameters, in this case β's. In general we call this an α-based contrast.

$$f = L_2[\alpha_1 + \tfrac{1}{3}(\beta_1 + \beta_2 + \beta_3)] + L_3[\alpha_2 + \tfrac{1}{1}(\quad + \beta_3)] + L_4[\alpha_3 + \tfrac{1}{2}(\beta_1 \quad + \beta_3)] + [-L_2 - L_3 - L_4][\alpha_4 + \tfrac{1}{3}(\beta_1 + \beta_2 + \beta_3)].$$

Algebraic simplification of this expression yields the coefficients shown in the output.

Examples: Coefficient of α_1 is L_2

Coefficient of β_1 is $\frac{1}{3}L_2 + \frac{1}{2}L_4 + (-L_2 - L_3 - L_4)\frac{1}{3} = -\frac{1}{3}L_3 + \frac{1}{6}L_4 = -0.3333L_3 + 0.1667L_4$.

Any $r = a - 1 = 3$ values of $[L_2 \;\; L_3 \;\; L_4]$ used in f such that the $r = 3$ resulting f_i's are linearly independent (LIN can be used. $R(\alpha|\mu)$ is then the numerator sum of squares for testing $H: f_i = 0$, $i = 1, 2, 3$.

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

TYPE I ESTIMABLE FUNCTIONS FOR: B

EFFECT		COEFFICIENTS
INTERCEPT		0
A	1X	0
	2Y	0
	3Z	0
	4W	0
B	A	L6
	B	L7
	C	-L6-L7

TYPE I, for B in the sequence A, B.

$R(\beta|\mu,\alpha)$ is the sum of squares explained.

LM 282 indicates that this tests

$$H: \beta_1 = \beta_2 = \beta_3 .$$

A contrast among these is

$$f = L_6\beta_1 + L_7\beta_2 + (-L_6 - L_7)\beta_3$$

which is the output.

Any two linearly independent (LIN) values given to $\ell' = [L_6 \quad L_7]$ then defines two values of f. f_1 and f_2 say, such that

$$H: f_i = 0 \quad i = 1, 2$$

is tested by the F-statistic using $R(\beta|\mu,\alpha)$ in the numerator.

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

TYPE II ESTIMABLE FUNCTIONS FOR: A

EFFECT		COEFFICIENTS
INTERCEPT		0
A	1X	L2
	2Y	L3
	3Z	L4
	4W	-L2-L3-L4
B	A	0
	B	0
	C	0

For TYPE II, the sums of squares explained are those for fitting each factor after all appropriate others; i.e., none that are crossed with or nested within the factor of interest.

Example: This output is for explaining $R(\alpha|\mu,\beta)$.

Analogous to the preceding page this tests

$$H: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 .$$

A general contrast among these is

$$f = L_2\alpha_1 + L_3\alpha_2 + L_4\alpha_3 + (-L_2 - L_3 - L_4)\alpha_4$$

which is the output.

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

TYPE III ESTIMABLE FUNCTIONS FOR: A

EFFECT		COEFFICIENTS
INTERCEPT		0
A	1X	L2
	2Y	L3
	3Z	L4
	4W	-L2-L3-L4
B	A	0
	B	0
	C	0

TYPES III and IV for this data set are the same as TYPE II. Detailed explanation is provided on the output of data set (3).

Sums of squares are annotated on p. 13.

UNBALANCED DATA, NO INTERACTION, N(I,J) = 0 OR 1
LINEAR MODELS BY S R SEARLE PAGE 262

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE	DF		SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	5		420.00000000	84.00000000	21.00	0.0153	0.972222	16.6667
ERROR	3	SSE	12.00000000	4.00000000		STD DEV		Y MEAN
CORRECTED TOTAL	8		432.00000000			2.00000000		12.00000000

SOURCE	DF		TYPE I SS	F VALUE	PR > F	DF		TYPE II SS	F VALUE	PR > F
A	3	$R(\alpha\|\mu)$	162.00000000	13.50	0.0301	3	$R(\alpha\|\mu,\beta)$	271.50000000	22.62	0.0146
B	2	$R(\beta\|\mu,\alpha)$	258.00000000	32.25	0.0094	2	$R(\beta\|\mu,\alpha)$	258.00000000	32.25	0.0094

SOURCE	DF	TYPE III SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
A	3	271.50000000	22.62	0.0146	3	271.50000000	22.62	0.0146
B	2	258.00000000	32.25	0.0094	2	258.00000000	32.25	0.0094

PARAMETER		ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT		17.00000000 B	11.13	0.0016	1.52752523
A	1X	9.00000000 B	5.51	0.0118	1.63299316
	2Y	-8.00000000 B	-3.18	0.0501	2.51661148
	3Z	-3.00000000 B	-1.57	0.2152	1.91485422
	4W	0.00000000 B	.	.	.
B	A	-10.00000000 B	-6.12	0.0088	1.63299316
	B	-14.00000000 B	-7.31	0.0053	1.91485422
	C	0.00000000 B	.	.	.

← This is valueless.

The solution is not unique; neither is its variance, standard error, or "t".

NOTE: AN INFINITE NUMBER OF SOLUTIONS TO THE NORMAL EQUATIONS EXIST. ESTIMATES FOLLOWED BY THE LETTER B ARE BIASED. REFER TO THE GENERAL FORM OF ESTIMABLE FUNCTIONS TO SEE WHAT THE EXPECTED VALUE OF THE BIASED ESTIMATORS ARE.

"Solution" is not explained.

This is a solution vector.

$$b^{o'} = [\underbrace{17}_{\mu}\ \underbrace{9\ \ -8\ \ -3\ \ 0}_{\alpha's}\ \underbrace{-10\ \ -14\ \ 0}_{\beta's}]$$

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

TYPE I ESTIMABLE FUNCTIONS FOR: B

EFFECT		COEFFICIENTS
INTERCEPT		0
B	A	L2
	B	L3
	C	-L2-L3
A	1X	0.0833*L2+0.25*L3
	2Y	-0.25*L2-0.25*L3
	3Z	0.0833*L2-0.25*L3
	4W	0.0833*L2+0.25*L3

TYPE I for B in the sequence B, A, AB.

$R(\beta|\mu)$ is explained here.

LM 282 indicates the hypothesis to be H: $\beta_j + \frac{1}{n_{\cdot j}} \sum_{i=1}^{a} \alpha_i$ equal.

n_{ij}'s			
1	1	1	3
–	–	1	1
1	-	1	2
1	1	1	3
3	2	4	9

$$H: \begin{Bmatrix} \beta_1 + \frac{1}{3}(\alpha_1 + \alpha_3 + \alpha_4) \\ \beta_2 + \frac{1}{2}(\alpha_1 + \alpha_4) \\ \beta_3 + \frac{1}{4}(\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4) \end{Bmatrix} \text{ equal.}$$

A general linear contrast of the β's plus a "mess" of α's is

$$f = L_2[\beta_1 + \tfrac{1}{3}(\alpha_1 + \alpha_3 + \alpha_4)] + L_3[\beta_2 + \tfrac{1}{2}(\alpha_1 + \alpha_4)] + (-L_2 - L_3)[\beta_3 + \tfrac{1}{4}(\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4)].$$

This yields the output.

Example: Coefficient of α_1: $\frac{1}{3}L_2 + \frac{1}{2}L_3 + (-L_2 - L_3)\frac{1}{4} = (\frac{1}{3} - \frac{1}{4})L_2 + (\frac{1}{2} - \frac{1}{4})L_3$

$= -\frac{1}{12}L_2 + \frac{1}{4}L_3$

$= -.0833L_2 + 0.25L_3$

Coefficient of α_2: $\frac{1}{2}L_3 + (-L_2 - L_3)\frac{1}{4} = \frac{1}{4}L_2 - \frac{1}{4}L_3$

$= .25L_2 - .25L_3$.

DATA SET 2
UNBALANCED DATA, NO INTERACTION, N(I,J) = 0 OR 1
LINEAR MODELS BY S R SEARLE PAGE 262

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	5	420.00000000	84.00000000	21.00	0.0153	0.972222	16.6667
ERROR	3	12.00000000	4.00000000		STD DEV		Y MEAN
CORRECTED TOTAL	8	432.00000000			2.00000000		12.00000000

SOURCE	DF		TYPE I SS	F VALUE	PR > F
B	2	$R(\beta\|\mu)$	148.50000000	18.56	0.0204
A	3	$R(\alpha\|\mu,\beta)$	271.50000000	22.63	0.0146

DATA SET 3
UNBALANCED DATA, TWO WAY CROSSED CLASSIFICATION
WITH INTERACTION N(I,J) > 0
EXAMPLE FROM BU-608-M BY S R SEARLE

SAS GLM Data Set 3

GENERAL LINEAR MODELS PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
A	2	1 2
B	3	1 2 3

NUMBER OF OBSERVATIONS IN DATA SET = 8

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

GENERAL FORM OF ESTIMABLE FUNCTIONS

EFFECT		COEFFICIENTS
INTERCEPT		L1
A	1	L2
	2	L1-L2
B	1	L4
	2	L5
	3	L1-L4-L5
A*B	1 1	L7
	1 2	L8
	1 3	L2-L7-L8
	2 1	L4-L7
	2 2	L5-L8
	2 3	L1-L2-L4-L5+L7+L8

For the general estimable function $\underset{\sim}{w}'\underset{\sim}{H}\underset{\sim}{b}$, the value of $\underset{\sim}{H} = \underset{\sim}{G}\underset{\sim}{X}'\underset{\sim}{X}$ in this example, with $\underset{\sim}{G} = \begin{bmatrix} \underset{\sim}{0} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{D} \end{bmatrix}$ for $\underset{\sim}{D}$ being $\underset{\sim}{D} = \mathrm{diag}\{1/n_{ij}\}$, is

$$\underset{\sim}{H} = \begin{bmatrix} \underset{\sim}{1}_3 & \underset{\sim}{1}_3 & \underset{\sim}{0} & \underset{\sim}{I}_3 & \underset{\sim}{I}_3 & \underset{\sim}{0} \\ \underset{\sim}{1}_3 & \underset{\sim}{0} & \underset{\sim}{1}_3 & \underset{\sim}{I}_3 & \underset{\sim}{0} & \underset{\sim}{I}_3 \end{bmatrix}.$$

Then

$$(\underset{\sim}{w}'\underset{\sim}{H})' = \begin{bmatrix} w_1 + w_2 + w_3 + w_4 + w_5 + w_6 \\ w_1 + w_2 + w_3 \\ w_4 + w_5 + w_6 \\ w_1 + w_4 \\ w_2 + w_5 \\ w_3 + w_6 \\ w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \\ w_6 \end{bmatrix}, \text{ and on defining L's in terms of w's this is } (\underset{\sim}{w}'\underset{\sim}{H})' = \begin{bmatrix} L_1 \\ L_2 \\ L_1 - L_2 \\ L_4 \\ L_5 \\ L_1 - L_4 - L_5 \\ L_7 \\ L_8 \\ L_2 - L_7 - L_8 \\ L_4 - L_7 \\ L_5 - L_8 \\ L_1 - L_2 - L_4 - L_5 + L_7 + L_8 \end{bmatrix}$$

of the output.

DEPENDENT VARIABLE: Y

TYPE I ESTIMABLE FUNCTIONS FOR: A

EFFECT		COEFFICIENTS
INTERCEPT		0
A	1	L2
	2	-L2
B	1	0.25*L2
	2	-0.25*L2
	3	0
A*B	1 1	0.5*L2
	1 2	0.25*L2
	1 3	0.25*L2
	2 1	-0.25*L2
	2 2	-0.5*L2
	2 3	-0.25*L2

Notation

γ_{ij} for interaction $(\alpha\beta)_{ij}$

This is a TYPE I. The input sequence of factors is A, B, AB.

This output explains $R(\alpha|\mu)$.

LM 307 gives H as $H: \alpha_i + \frac{1}{n_{i.}} \sum_{j=1}^{b} n_{ij}(\beta_j + \gamma_{ij})$ all equal.

n_{ij}'s			
2	1	1	4
1	2	1	4
3	3	2	8

$$H: \left\{ \begin{matrix} \alpha_1 + \frac{1}{4}(2\beta_1 + \beta_2 + \beta_3 + 2\gamma_{11} + \gamma_{12} + \gamma_{13}) \\ \alpha_2 + \frac{1}{4}(\beta_1 + 2\beta_2 + \beta_3 + \gamma_{21} + 2\gamma_{22} + \gamma_{23}) \end{matrix} \right\} \text{ equal.}$$

A general α-based contrast of the α's, plus a "mess" of other parameters is

$$f = L_2[\alpha_1 + \tfrac{1}{4}(2\beta_1 + \beta_2 + \beta_3 + 2\gamma_{11} + \gamma_{12} + \gamma_{13})] + (-L_2)[\alpha_2 + \tfrac{1}{4}(\beta_1 + 2\beta_2 + \beta_3 + \gamma_{21} + 2\gamma_{22} + \gamma_{23})] = 0.$$

This is the output.

Examples of coefficients

of β_1 : $\frac{1}{2}L_2 - \frac{1}{4}L_2 = \frac{1}{4}L_2 = .25L_2$

of γ_{13}: $\frac{1}{4}L_2 = .25L_2$.

DEPENDENT VARIABLE: Y

TYPE I ESTIMABLE FUNCTIONS FOR: B

EFFECT		COEFFICIENTS
INTERCEPT		0
A	1	0
	2	0
B	1	L4
	2	L5
	3	-L4-L5
A*B	1 1	0.6061*L4+0.0606*L5
	1 2	-0.0606*L4+0.3939*L5
	1 3	-0.5455*L4-0.4545*L5
	2 1	0.3939*L4-0.0606*L5
	2 2	0.0606*L4+0.6061*L5
	2 3	-0.4545*L4-0.5455*L5

TYPE I, for B in sequence A, B, AB.

$R(\beta|\mu,\alpha)$ is the sum of squares explained.

From LM 308 the hypothesis is H: $\psi_j = 0$ for $j = 1 \cdots b-1$

for, from (87), LM 305

$$\psi_j = \left(n_{.j} - \sum_{j=1}^{a} \frac{n_{ij}^2}{n_{i.}}\right)\beta_j - \sum_{j'\neq j}^{b}\sum_{i=1}^{a}\frac{n_{ij}n_{ij'}}{n_{i.}}\beta_j + \sum_i\left(n_{ij} - \frac{n_{ij}^2}{n_{i.}}\right)\gamma_{ij} - \sum_{j'\neq j}^{b}\left(\sum_{i=1}^{a}\frac{n_{ij}n_{ij'}}{n_{i.}}\right)\gamma_{ij'} \; . \qquad (1)$$

Define this as

$$\psi_j = \sum_{j'=1}^{b} c_{jj'}\beta_{j'} + \sum_i\sum_{j'}\lambda_{j,ij'}\gamma_{ij'}, \quad \text{with } c_{j'j} = c_{jj'} = \delta_{jj'}n_{.j} - \sum_{i=1}^{a} n_{ij}n_{ij'}/n_{i.}$$

$$\lambda_{j,ij'} = \delta_{jj'}n_{ij} - n_{ij}n_{ij'}/n_{i.} \; . \qquad (\text{Kronecker } \delta_{jj'}) \qquad (2)$$

Then the $c_{jj'}$'s are those of (17), LM 267, and we have the following equalities. For each ψ_j:

$$(\Sigma \text{ coefficients of all } \beta_j\text{'s}) = 0 \; ; \text{ i.e., } \sum_{j'=1}^{b} c_{jj'} = 0 \; . \; \forall\, j \qquad (3)$$

$$\text{For every } i: \; (\sum_{j'} \text{ coefficients of } \gamma_{ij'}) = 0 \; ; \text{ i.e., } \sum_{j'=1}^{b} \lambda_{j\,ij'} = 0 \; . \; \forall\, i,j \qquad (4)$$

$$\text{For every } j': \; (\sum_{i} \text{ coefficients of } \gamma_{ij'}) = c_{ij'} ; \text{ i.e., } \sum_{i=1}^{a} \lambda_{j,ij'} = c_{jj'} . \; \forall\, jj' \qquad (5)$$

All of the c's and λ's are known values. But equation (3) means that b - 1 of the c's determines the other and, for filled cells, (4) and (5) mean that (a-1)(b-1) of the λ_j's determine the others. And these statements are true b - 1 times, for b - 1 ψ_j's – since $\sum_{j=1}^{b}\psi_j = 0$ – see LM 308. These equalities are the basis for development of the program output. This is now illustrated for the example: a = 2, b = 3.

Coefficient of

In	β_1	β_2	β_3	γ_{11}	γ_{12}	γ_{13}	γ_{21}	γ_{22}	γ_{23}
ψ_1:	c_{11}	c_{12}	$-c_{11} - c_{12}$	$\lambda_{1,11}$	$\lambda_{1,12}$	$-\lambda_{1,11} - \lambda_{1,12}$	$c_{11} - \lambda_{1,11}$	$c_{12} - \lambda_{1,12}$	$-c_{11} - c_{12} + \lambda_{1,11} + \lambda_{1,12}$
ψ_2:	c_{21}	c_{22}	$-c_{21} - c_{22}$	$\lambda_{2,11}$	$\lambda_{2,12}$	$-\lambda_{2,11} - \lambda_{2,12}$	$c_{21} - \lambda_{2,11}$	$c_{22} - \lambda_{2,12}$	$-c_{21} - c_{22} + \lambda_{2,11} + \lambda_{2,12}$
			from (3)			from (4)	from (5)	from (5)	from (3), (4) and (5)

All these values are known – from (2). But from this table it is clear that they are all determined by just the values

$$\begin{matrix} c_{11} & c_{12} & \lambda_{1,11} & \lambda_{1,12} \\ c_{21} & c_{22} & \lambda_{2,11} & \lambda_{2,12} \end{matrix}.$$

Thus

$$\underbrace{\begin{bmatrix} \psi_1 \\ \psi_2 \end{bmatrix}}_{(b-1)\times 1} = \begin{bmatrix} \underbrace{\begin{matrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{matrix}}_{(b-1)\times(b-1)} & \underbrace{\begin{matrix} \lambda_{1,11} & \lambda_{1,12} \\ \lambda_{2,11} & \lambda_{2,12} \end{matrix}}_{(b-1)\times \text{ rank of interaction}} \end{bmatrix} \underbrace{\begin{bmatrix} 1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 & 0 & -1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & -1 & 1 \end{bmatrix}}_{\text{Define it as } \underset{\sim}{M}} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \gamma_{11} \\ \gamma_{12} \\ \gamma_{13} \\ \gamma_{21} \\ \gamma_{22} \\ \gamma_{23} \end{bmatrix} \qquad (6)$$

ψ_1 and ψ_2 are contrasts – from (3) and (4). Therefore any linear combinations of ψ_1 and ψ_2 is a contrast (because any linear combination of contrasts is a contrast).

Consider $p_1\psi_1 + p_2\psi_2$ as any linear contrast. From (6),

$$p_1\psi_1 + p_2\psi_2 = [p_1 \quad p_2] \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \left[\begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix} \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}^{-1} \begin{pmatrix} \lambda_{1,11} & \lambda_{1\ 12} \\ \lambda_{2,11} & \lambda_{2,12} \end{pmatrix} \right] \underset{\sim}{M} \begin{bmatrix} \underset{\sim}{\beta} \\ \underset{\sim}{\gamma} \end{bmatrix}.$$

Since p_1 and p_2 are arbitrary, write

$$L_4 = p_1c_{11} + p_2c_2 \quad \text{and} \quad L_5 = p_1c_{12} + p_2c_{22}$$

and then an arbitrary contrast is

$$f = [L_4 \quad L_5] \left[\begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix} \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}^{-1} \begin{pmatrix} \lambda_{1,11} & \lambda_{1,12} \\ \lambda_{2,11} & \lambda_{2,12} \end{pmatrix} \right] \underset{\sim}{M} \begin{bmatrix} \underset{\sim}{\beta} \\ \underset{\sim}{\gamma} \end{bmatrix}. \qquad (7)$$

Since the c's and λ's are available numbers, from (2), it is (7) that determines the output.

n_{ij}'s			
2	1	1	4
1	2	1	4
3	3	2	8

$$c_{11} = 3 - \left(\frac{4}{4} + \frac{1}{4}\right) = \frac{7}{4} \qquad c_{12} = -\frac{2(1)}{4} - \frac{1(2)}{4} = -1$$

$$c_{22} = 3 - \left(\frac{1}{4} + \frac{4}{4}\right) = \frac{7}{4}$$

$$\lambda_{1,11} = 2 - \frac{4}{4} = 1 \qquad \lambda_{1,12} = -\frac{2(1)}{4} = -\frac{1}{2}$$

$$\lambda_{2,11} = -\frac{2(1)}{4} = -\frac{1}{2} \qquad \lambda_{2,12} = 1 - \frac{1}{4} = \frac{3}{4}$$

$$\begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}^{-1} \begin{bmatrix} \lambda_{1,11} & \lambda_{1,12} \\ \lambda_{2,11} & \lambda_{2,12} \end{bmatrix} = \begin{bmatrix} \frac{7}{4} & -1 \\ -1 & \frac{7}{4} \end{bmatrix}^{-1} \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & \frac{3}{4} \end{bmatrix} = \frac{1}{99} \begin{bmatrix} 60 & -6 \\ 6 & 39 \end{bmatrix} = \begin{bmatrix} .6060 & -.0606 \\ .0606 & .3939 \end{bmatrix}.$$

Therefore the arbitrary contrast is

$$f = [L_4 \quad L_5] \begin{bmatrix} 1 & 0 & .6060 & -.0606 \\ 0 & 1 & .0606 & .3939 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 & 0 & -1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} \underset{\sim}{\beta} \\ \underset{\sim}{\gamma} \end{bmatrix}$$

$$= [L_4 \quad L_5] \begin{bmatrix} 1 & 0 & -1 & .6060 & -.0606 & -.5455 & .3939 & .0606 & -.4545 \\ 0 & 1 & -1 & .0606 & .3939 & -.4545 & -.0606 & .6061 & .5455 \end{bmatrix} \begin{bmatrix} \underset{\sim}{\beta} \\ \underset{\sim}{\gamma} \end{bmatrix},$$

which leads at once to the computer output.

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

TYPE I ESTIMABLE FUNCTIONS FOR: A*B

EFFECT		COEFFICIENTS
INTERCEPT		0
A	1	0
	2	0
B	1	0
	2	0
	3	0
A*B	1 1	L7
	1 2	L8
	1 3	-L7-L8
	2 1	-L7
	2 2	-L8
	2 3	L7+L8

TYPE I for AB in sequence A B, AB.

$R(\gamma|\mu,\alpha,\beta)$ is the sum of squares explained.

Tests (LM 310) s - a - b + 1 = 6 - 2 - 3 + 1 = 2 linearly independent contrasts

$$H: \gamma_{ij} - \gamma_{ij'} - \gamma_{i'j} + \gamma_{i'j'} = 0 .$$

The pattern of filled cells is

✓	✓	✓
✓	✓	✓

.

Two possibilities are

$$H: \begin{cases} \gamma_{11} - \gamma_{12} - \gamma_{21} + \gamma_{22} = 0 \\ \gamma_{12} - \gamma_{13} - \gamma_{22} + \gamma_{23} = 0 \end{cases} .$$

Any linear combination of these is a contrast and its form is

$$f = L_7(\gamma_{11} - \gamma_{12} - \gamma_{21} + \gamma_{22}) + (L_7 + L_8)(\gamma_{12} - \gamma_{13} - \gamma_{22} + \gamma_{23})$$

$$= L_7\gamma_{11} + L_8\gamma_{12} - (L_7 + L_8)\gamma_{13} - L_7\gamma_{21} - L_8\gamma_{22} + (L_7 + L_8)\gamma_{23} ,$$

which is the output.

DEPENDENT VARIABLE: Y

TYPE II ESTIMABLE FUNCTIONS FOR: A

EFFECT		COEFFICIENTS
INTERCEPT		0
A	1	L2
	2	-L2
B	1	0
	2	0
	3	0
A*B	1 1	0.3636*L2
	1 2	0.3636*L2
	1 3	0.2727*L2
	2 1	-0.3636*L2
	2 2	-0.3636*L2
	2 3	-0.2727*L2

TYPE II explains R(each factor|all appropriate others).

In this case it is $R(\alpha|\mu,\beta)$.

Use procedure analogous to $R(\beta|\mu,\alpha)$, p. 26 et seq.

Comparable to (2) we have, from LM 308,

$$\varphi_i = \sum_{i'=1}^{a} t_{ii'}\alpha_{i'} + \sum_j\sum_{i'} \theta_{i,i'j}\gamma_{i'j} \quad \text{with} \quad t_{ii'} = \delta_{ii'}n_{i.} - \sum_{j=1}^{b} n_{ij}n_{i'j}/n_{.j}$$

$$\theta_{i,i'j} = \delta_{ii'}n_{ij} - n_{ij}n_{i'j}/n_{.j}$$

and, like (3), (4) and (5)

$$\sum_{i'=1}^{a} t_{ii'} = 0 \;\forall\; i, \qquad \sum_{i'=1}^{a} \theta_{i,i'j} = 0 \;\forall\; ij, \quad \text{and} \quad \sum_{j=1}^{b} \theta_{i,i'j} = t_{ii'} .$$

For these data, the arbitrary contrast (only one, because i = 1, 2) comparable to (7) is

	α_1	α_2	γ_{11}	γ_{12}	γ_{13}	γ_{21}	γ_{22}	γ_{23}
$\nu_1 =$	t_{11}	$-t_{11}$	$\theta_{1,11}$	$\theta_{1,12}$	$t_{11}-\theta_{1,11}-\theta_{1,12}$	$-\theta_{1,11}$	$-\theta_{1,12}$	$-t_{11}+\theta_{1,11}+\theta_{1,12}$

$$= [t_{11} \quad \theta_{1,11} \quad \theta_{1,12}] \begin{bmatrix} 1 & -1 & 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & -1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 1 & -1 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} \underset{\sim}{\alpha} \\ \underset{\sim}{\gamma} \end{bmatrix}$$

n_{ij}'s			
2	1	1	4
1	2	1	4
3	3	2	8

$$t_{11} = 4 - \left(\frac{4}{3} + \frac{1}{3} + \frac{1}{2}\right) = \frac{11}{6} \qquad \theta_{1,11} = 2 - \frac{4}{3} = \frac{2}{3}$$

$$\theta_{1,12} = 1 - \frac{1}{3} = \frac{1}{3}$$

Hence an arbitrary contrast $(L_2/t_{11})\nu_1$ is

$$f = L_2\left[1 \quad \frac{6}{11}\left(\frac{2}{3}\right) \quad \frac{6}{11}\left(\frac{2}{3}\right)\right]\underset{\sim}{M}\begin{bmatrix} \underset{\sim}{\alpha} \\ \underset{\sim}{\gamma} \end{bmatrix} = L_2\left[1 \quad \frac{4}{11} \quad \frac{4}{11}\right]\begin{bmatrix} 1 & -1 & 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & -1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 1 & -1 & 0 & -1 & 1 \end{bmatrix}\begin{bmatrix} \underset{\sim}{\alpha} \\ \underset{\sim}{\gamma} \end{bmatrix}$$ which yields the output.

UNBALANCED DATA, TWO WAY CROSSED CLASSIFICATION
WITH INTERACTION N(I,J) > 0
EXAMPLE FROM BU-608-M BY S R SEARLE

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

TYPE II ESTIMABLE FUNCTIONS FOR: B

TYPE II for B, in a model of A, B, AB;
i.e., $R(\beta|\mu,\alpha)$. See output p. 26.

EFFECT		COEFFICIENTS
INTERCEPT		0
A	1	0
	2	0
B	1	L4
	2	L5
	3	-L4-L5
A*B	1 1	0.6061*L4+0.0606*L5
	1 2	-0.0606*L4+0.3939*L5
	1 3	-0.5455*L4-0.4545*L5
	2 1	0.3939*L4-0.0606*L5
	2 2	0.0606*L4+0.6061*L5
	2 3	-0.4545*L4-0.5455*L5

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

TYPE III ESTIMABLE FUNCTIONS FOR: A

EFFECT		COEFFICIENTS
INTERCEPT		0
A	1	L2
	2	-L2
B	1	0
	2	0
	3	0
A*B	1 1	0.3333*L2
	1 2	0.3333*L2
	1 3	0.3333*L2
	2 1	-0.3333*L2
	2 2	-0.3333*L2
	2 3	-0.3333*L2

TYPE III: the "usual constraints" are utilized,

$$\Sigma\alpha_i = 0 \qquad \Sigma\beta_j = 0 \qquad \sum_i \gamma_{ij} = 0 \ \forall\ j, \qquad \Sigma\gamma_{ij} = 0 \ \forall\ j\ .$$

They are used in a special, computational manner. Pages 23-25 of BU-608-M describes how they are used. The result is what is called $R^*(\alpha|\mu,\alpha,\gamma)\Sigma$. For the <u>all</u> <u>cells</u> <u>filled</u> case this is identical to SSA_w, the sum of squares due to rows in the weighted squares of means analysis, LM 369-372.

The hypothesis tested (LM 371) is

$$H:\ \alpha_i + \bar{\gamma}_{i.} \text{ all equal.}$$

This is equivalent to the contrast

$$H:\ \alpha_i + \bar{\gamma}_{i.} - \alpha_{i'} - \bar{\gamma}_{i'.} = 0$$

for a - 1 linearly independent such contrasts. In data set 3, a = 2 so we can use any arbitrary multiple of H:

$$f = L_2(\alpha_1 + \bar{\gamma}_{1.} - \alpha_2 - \bar{\gamma}_{2.}) \equiv L_2[\alpha_1 + \tfrac{1}{3}(\gamma_{11} + \gamma_{12} + \gamma_{13}) - \alpha_2 - \tfrac{1}{3}(\gamma_{21} + \gamma_{22} + \gamma_{23})],$$

which is the output.

UNBALANCED DATA, TWO WAY CROSSED CLASSIFICATION
WITH INTERACTION N(I,J) > 0
EXAMPLE FROM BU-608-M BY S R SEARLE

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

TYPE III ESTIMABLE FUNCTIONS FOR: B

EFFECT		COEFFICIENTS
INTERCEPT		0
A	1	0
	2	0
B	1	L4
	2	L5
	3	-L4-L5
A*B	1 1	0.5*L4
	1 2	0.5*L5
	1 3	-0.5*L4-0.5*L5
	2 1	0.5*L4
	2 2	0.5*L5
	2 3	-0.5*L4-0.5*L5

TYPE III explains $R^*(\beta|\mu,\alpha,\gamma) = SSB_w$, for all cells filled.

The hypothesis is H: $\beta_j + \bar{\gamma}_{.j}$ all equal;

i.e., H: $\beta_1 + \bar{\gamma}_{.1} = \beta_2 + \bar{\gamma}_{.2} = \beta_3 + \bar{\gamma}_{.3}$.

This is equivalent to

$$H: \beta_1 + \bar{\gamma}_{.1} - (\beta_3 + \bar{\gamma}_{.3}) = 0$$
$$\beta_2 + \bar{\gamma}_{.2} - (\beta_3 + \bar{\gamma}_{.3}) = 0 .$$

An arbitrary combination is

$$f = L_4[\beta_1 + \bar{\gamma}_{.1} - (\beta_3 + \bar{\gamma}_{.3})] + L_5[\beta_2 + \bar{\gamma}_{.2} - (\beta_3 + \bar{\gamma}_{.3})],$$

which is the output:

e.g., term in β_3 is $\beta_3(-L_4 - L_5)$

term in γ_{21} is $(\frac{1}{2}L_4)$.

UNBALANCED DATA. TWO WAY CROSSED CLASSIFICATION
WITH INTERACTION N(I,J) > 0
EXAMPLE FROM BU-608-M BY S R SEARLE

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

TYPE IV ESTIMABLE FUNCTIONS FOR: A

EFFECT		COEFFICIENTS
INTERCEPT		0
A	1	L2
	2	-L2
B	1	0
	2	0
	3	0
A*B	1 1	0.3333*L2
	1 2	0.3333*L2
	1 3	0.3333*L2
	2 1	-0.3333*L2
	2 2	-0.3333*L2
	2 3	-0.3333*L2

TYPE IV sums of squares for this data set are the same as TYPE III. Detailed explanation is provided on the output of data set (4).

UNBALANCED DATA, TWO WAY CROSSED CLASSIFICATION
WITH INTERACTION N(I,J) > 0
EXAMPLE FROM BU-608-M BY S R SEARLE

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE	DF		SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	5	SSR_m	56.00000000	11.20000000	2.24	0.3369	0.848485	31.9438
ERROR	2	SSE	10.00000000	5.00000000		STD DEV		Y MEAN
CORRECTED TOTAL	7	SST_m	66.00000000			2.23606798		7.00000000

SOURCE	DF		TYPE I SS	F VALUE	PR > F	DF		TYPE II SS	F VALUE	PR > F
A	1	$R(\alpha\|\mu)$	8.00000000	1.60	0.3333	1	$R(\alpha\|\mu,\beta)$	13.63636364	2.73	0.2404
B	2	$R(\beta\|\mu,\alpha)$	11.63636364	1.16	0.4622	2	$R(\beta\|\mu,\alpha)$	11.63636364	1.16	0.4622
A*B	2	$R(\gamma\|\mu,\alpha,\beta)$	36.36363636	3.64	0.2157	2	$R(\gamma\|\mu,\alpha,\beta)$	36.36363636	3.64	0.2157

SOURCE	DF		TYPE III SS	F VALUE	PR > F	DF		TYPE IV SS	F VALUE	PR > F
A	1	SSA_w	20.00000000	4.00	0.1835	1	Same as III	20.00000000	4.00	0.1835
B	2	SSB_w	5.33333333	0.53	0.6522	2		5.33333333	0.53	0.6522
A*B	2		36.36363636	3.64	0.2157	2		36.36363636	3.64	0.2157

PARAMETER		ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT		12.00000000 B	5.37	0.0330	2.23606798
A	1	-10.00000000 B	-3.16	0.0871	3.16227766
	2	0.00000000 B	.	.	.
B	1	-4.00000000 B	-1.26	0.3333	3.16227766
	2	-6.00000000 B	-2.19	0.1598	2.73861279
	3	0.00000000 B	.	.	.
A*B	1 1	10.00000000 B	2.39	0.1393	4.18330013
	1 2	10.00000000 B	2.39	0.1393	4.18330013
	1 3	0.00000000 B	.	.	.
	2 1	0.00000000 B	.	.	.
	2 2	0.00000000 B	.	.	.
	2 3	0.00000000 B	.	.	.

Comments similar to those on Data Set 2 apply here.

NOTE: AN INFINITE NUMBER OF SOLUTIONS TO THE NORMAL EQUATIONS EXIST. ESTIMATES FOLLOWED BY THE LETTER B ARE BIASED. REFER TO THE GENERAL FORM OF ESTIMABLE FUNCTIONS TO SEE WHAT THE EXPECTED VALUE OF THE BIASED ESTIMATORS ARE.

UNBALANCED DATA, TWO WAY CROSSED CLASSIFICATION
WITH INTERACTION N(I,J) > 0
EXAMPLE FROM BU-608-M BY S R SEARLE

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

TYPE I ESTIMABLE FUNCTIONS FOR: B

EFFECT		COEFFICIENTS
INTERCEPT		0
B	1	L2
	2	L3
	3	-L2-L3
A	1	0.1667*L2-0.1667*L3
	2	-0.1667*L2+0.1667*L3
A*B	1 1	0.6667*L2
	1 2	0.3333*L3
	1 3	-0.5*L2-0.5*L3
	2 1	0.3333*L2
	2 2	0.6667*L3
	2 3	-0.5*L2-0.5*L3

TYPE I for B in the sequence B, A, AB.

The output explains $R(\beta|\mu)$.

The hypothesis tested by this is (LM 308)

$$H: \beta_j + \frac{1}{n_{.j}} \Sigma n_{ij}(\alpha_i + \gamma_{ij}) \text{ all equal}$$

n_{ij}'s			
2	1	1	4
1	2	1	4
3	3	2	8

$$H: \left\{\begin{matrix} \beta_1 + \frac{1}{3}(2\alpha_1 + \alpha_2 + 2\gamma_{11} + \gamma_{21}) \\ \beta_2 + \frac{1}{3}(\alpha_1 + 2\alpha_2 + \gamma_{12} + \gamma_{22}) \\ \beta_3 + \frac{1}{2}(\alpha_1 + \alpha_2 + \gamma_{13} + \gamma_{23}) \end{matrix}\right\} \text{ equal}$$

A general contrast of the β's, plus a mess of other parameters is

$$f = L_2[\beta_1 + \tfrac{1}{3}(2\alpha_1 + \alpha_2 + 2\gamma_{11} + \gamma_{21})] + L_3[\beta_2 + \tfrac{1}{3}(\alpha_1 + 2\alpha_2 + \gamma_{12} + \gamma_{22})]$$
$$+ (-L_2 - L_3)[\beta_3 + \tfrac{1}{2}(\alpha_1 + \alpha_2 + \gamma_{13} + \gamma_{23})],$$

which is the output.

Example : term in α_1 is $\frac{2}{3}L_2 + \frac{1}{3}L_3 + (-L_2 - L_3)\frac{1}{2} = \frac{1}{6}L_2 - \frac{1}{6}L_3$.

DATA SET 3
UNBALANCED DATA, TWO WAY CROSSED CLASSIFICATION
WITH INTERACTION N(I,J) > 0
EXAMPLE FROM BU-608-M BY S R SEARLE

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	5	56.00000000	11.20000000	2.24	0.3369	0.848485	31.9438
ERROR	2	10.00000000	5.00000000		STD DEV		Y MEAN
CORRECTED TOTAL	7	66.00000000			2.23606798		7.00000000

SOURCE	DF		TYPE I SS	F VALUE	PR > F
B	2	$R(\beta\|\mu)$	6.00000000	0.60	0.6250
A	1	$R(\alpha\|\mu,\beta)$	13.63636364	2.73	0.2404
A*B	2	$R(\gamma\|\mu,\alpha,\beta)$	36.36363636	3.64	0.2157

DATA SET 4
UNBALANCED DATA, TWO WAY CROSSED CLASSIFICATION
WITH INTERACTION SOME N(I,J) = 0
EXAMPLE ADAPTED FROM BU-417-M BY S R SEARLE

GENERAL LINEAR MODELS PROCEDURE

CLASS LEVEL INFORMATION

CLASS	LEVELS	VALUES
A	2	1 2
B	3	1 2 3

SAS GLM Data Set 4

NUMBER OF OBSERVATIONS IN DATA SET = 10

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

TYPE I ESTIMABLE FUNCTIONS FOR: A

EFFECT		COEFFICIENTS
INTERCEPT		0
A	1	L2
	2	-L2
B	1	0
	2	-0.1667*L2
	3	0.1667*L2
A*B	1 1	0.5*L2
	1 2	0.3333*L2
	1 3	0.1667*L2
	2 1	-0.5*L2
	2 2	-0.5*L2

TYPE I for A, in the sequence A, B, AB.

Explains $R(\alpha|\mu)$.

This tests

$$H: \alpha_i + \frac{1}{n_{i\cdot}} \sum_j n_{ij}(\beta_j + \gamma_{ij}) \text{ all equal.}$$

n_{ij}'s			
3	2	1	6
2	2	-	4
5	4	1	10

$$H: \begin{matrix} \alpha_1 + \frac{1}{6}(3\beta_1 + 2\beta_2 + \beta_3) + \frac{1}{6}(3\gamma_{11} + 2\gamma_{12} + \gamma_{13}) \\ \alpha_2 + \frac{1}{4}(2\beta_1 + 2\beta_2) \quad + \frac{1}{4}(2\gamma_{21} + 2\gamma_{22}) \end{matrix} \text{ equal.}$$

A general α-based contrast among the α's plus a "mess" of other parameters is

$$f = L_2[\alpha_1 + \tfrac{1}{6}(3\beta_1 + 2\beta_2 + \beta_3) + \tfrac{1}{6}(3\gamma_{11} + 2\gamma_{12} + \gamma_{13})]$$
$$- L_2[\alpha_2 + \tfrac{1}{4}(2\beta_1 + \beta_2) + \tfrac{1}{4}(2\gamma_{21} + \gamma_{22})] .$$

This yields the output; e.g.,

term in β_2	$\frac{2}{6}L_2 - \frac{2}{4}L_2$	$= -\frac{1}{6}L_2 =$	$-0.1667L_2$
term in γ_{12}	$\frac{2}{6}L_2$	$= \frac{1}{3}L_2 =$	$0.3333L_2$
term in γ_{13}	$\frac{1}{6}L_2$	$= \frac{1}{6}L_2 =$	$0.1667L_2$.

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

TYPE I ESTIMABLE FUNCTIONS FOR: B

EFFECT		COEFFICIENTS
INTERCEPT		0
A	1	0
	2	0
B	1	L4
	2	L5
	3	-L4-L5
A*B	1 1	0.8182*L4+0.2727*L5
	1 2	0.1818*L4+0.7273*L5
	1 3	-L4-L5
	2 1	0.1818*L4-0.2727*L5
	2 2	-0.1818*L4+0.2727*L5

TYPE I for B in the sequence A, B, AB.

Explains $R(\beta|\mu,\alpha)$

n_{ij}'s			
3	2	1	6
2	2	-	4
5	4	1	10

Se output p. 26, Data Set 3.

	β_1	β_2	β_3	γ_{11}	γ_{12}	γ_{13}	γ_{21}	γ_{22}
ψ_1	c_{11}	c_{12}	$-c_{11}-c_{12}$	λ_1	$c_{11}+c_{12}-\lambda_1$	$-c_{11}-c_{12}$	$c_{11}-\lambda_1$	$-c_{11}+\lambda_1$
ψ_2	c_{21}	c_{22}	$-c_{21}-c_{22}$	λ_2	$c_{21}+c_{22}-\lambda_1$	$-c_{21}-c_{22}$	$c_{12}-\lambda_2$	$-c_{12}+\lambda_2$

$$f = [L_4 \quad L_5]\begin{bmatrix} 1 & 0 & \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}^{-1} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} \\ 0 & 1 & \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 & 0 & 1 & -1 & 1 & -1 \\ 0 & 1 & -1 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} \underset{\sim}{\beta} \\ \underset{\sim}{\gamma} \end{bmatrix}$$

$$c_{11} = 5 - \left(\frac{9}{6} + \frac{4}{4}\right) = 2\tfrac{1}{2} \qquad c_{12} = -\left(\frac{6}{6} + \frac{4}{4}\right) = -2 \qquad \lambda_1 = 3 - \frac{9}{6} = 1\tfrac{1}{2}$$

$$c_{22} = 4 - \left(\frac{4}{6} + \frac{4}{4}\right) = 2\tfrac{1}{3} \qquad \lambda_2 = -\frac{6}{6} = -1$$

$$\begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}^{-1} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} 2\tfrac{1}{2} & -2 \\ -2 & 2\tfrac{1}{3} \end{bmatrix}^{-1} \begin{bmatrix} 1\tfrac{1}{2} \\ -1 \end{bmatrix} = \begin{bmatrix} 9/11 \\ 3/11 \end{bmatrix}$$

$$f = [L_4 \quad L_5]\begin{bmatrix} 1 & 0 & 9/11 \\ 0 & 1 & 3/11 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 & 0 & 1 & -1 & 1 & -1 \\ 0 & 1 & -1 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} \underset{\sim}{\beta} \\ \underset{\sim}{\gamma} \end{bmatrix} = [L_4 \quad L_5] \begin{bmatrix} 1 & 0 & -1 & 9/11 & 2/11 & -1 & 2/11 & -2/11 \\ 0 & 1 & -1 & 3/11 & 8/11 & -1 & -3/11 & 3/11 \end{bmatrix} \begin{bmatrix} \underset{\sim}{\beta} \\ \underset{\sim}{\gamma} \end{bmatrix},$$

which is the output.

UNBALANCED DATA, TWO WAY CROSSED CLASSIFICATION
WITH INTERACTION SOME N(I,J) = 0
EXAMPLE ADAPTED FROM BU-417-M BY S R SEARLE

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

TYPE I ESTIMABLE FUNCTIONS FOR: A*B

EFFECT		COEFFICIENTS
INTERCEPT		0
A	1	0
	2	0
B	1	0
	2	0
	3	0
A*B	1 1	L7
	1 2	-L7
	1 3	0
	2 1	-L7
	2 2	L7

TYPE I for AB in the sequence A, B, AB.

Explains $R(\gamma|\mu,\alpha,\beta)$.

Tests s - a - b + 1 linearly independent contrasts

$$H: \gamma_{ij} - \gamma_{ij'} - \gamma_{i'j} + \gamma_{i'j'} = 0 .$$

Since s - a - b + 1 = 5 - 2 - 3 + 1 = 1, there is only 1 contrast.

The pattern of filled cells is

✓	✓	✓
✓	✓	

.

The only contrast is

$$H: \gamma_{11} - \gamma_{12} - \gamma_{21} + \gamma_{22} = 0$$

A general form is

$$f = L_7(\gamma_{11} - \gamma_{12} - \gamma_{21} + \gamma_{22}) ,$$

which is the output.

DEPENDENT VARIABLE: Y

TYPE II ESTIMABLE FUNCTIONS FOR: A

EFFECT		COEFFICIENTS
INTERCEPT		0
A	1	L2
	2	-L2
B	1	0
	2	0
	3	0
A*B	1 1	0.5455*L2
	1 2	0.4545*L2
	1 3	0
	2 1	-0.5455*L2
	2 2	-0.4545*L2

TYPE II for A in the sequence A, B, AB.

Explains $R(\alpha|\mu,\beta)$

n_{ij}'s			
3	2	1	6
2	2	-	4
5	4	1	10

See output pp. 26 and 28, Data Set 3.

	α_1	α_2	γ_{11}	γ_{12}	γ_{13}	γ_{21}	γ_{22}
φ_1	t_{11}	$-t_{11}$	θ_1	$t_{11}-\theta_1$	0	$-\theta_1$	$-t_{11}+\theta_1$

$$t_{11} = 6 - \left(\frac{9}{5}+\frac{4}{4}+\frac{1}{1}\right) = 2.2 \qquad \theta_1 = 3 - \frac{3^2}{5} = 1.2 .$$

Then an arbitrary contrast is

$$f = L_2\begin{bmatrix}1 & \frac{\theta_1}{t_{11}}\end{bmatrix}\begin{bmatrix}1 & -1 & 0 & 1 & 0 & 0 & -1\\ 0 & 0 & 1 & -1 & 0 & -1 & 1\end{bmatrix}\begin{bmatrix}\underset{\sim}{\alpha}\\ \underset{\sim}{\gamma}\end{bmatrix}$$

and with $\theta_1/t_{11} = 12/22 = 6/11$.

$$f = L_2[1 \quad -1 \quad 6/11 \quad 5/11 \quad 0 \quad -6/11 \quad -5/11]\begin{bmatrix}\underset{\sim}{\alpha}\\ \underset{\sim}{\gamma}\end{bmatrix}$$

$$= L_2[1 \quad -1 \quad .5455 \quad .4545 \quad 0 \quad -.5455 \quad -.4545]\begin{bmatrix}\underset{\sim}{\alpha}\\ \underset{\sim}{\gamma}\end{bmatrix}.$$

This is the output.

DEPENDENT VARIABLE: Y

TYPE III ESTIMABLE FUNCTIONS FOR: A

EFFECT		COEFFICIENTS
INTERCEPT		0
A	1	L2
	2	-L2
B	1	0
	2	0
	3	0
A*B	1 1	0.5*L2
	1 2	0.5*L2
	1 3	0
	2 1	-0.5*L2
	2 2	-0.5*L2

TYPE III for A in the sequence A, B AB. $R^*(\alpha|\mu,\beta,\gamma)_\Sigma$ is explained.

When all cells are filled, $R^*(\alpha|\mu,\beta,\gamma)_\Sigma = SSA_w$ – see Data Set 3, output p. 31.

When some cells are empty, it explains a sum of squares suitable for testing a hypothesis that depends upon <u>which</u> cells have data but not upon how many observations in those cells.

Furthermore, it tests hypotheses about contrasts of α's and a "mess" of γ's – but <u>no</u> β's.

Finally, it is orthogonal to the AB contrasts of $R(\gamma|\mu,\alpha,\beta)$.

Pattern of filled cells:

✓	✓	✓
✓	✓	

With 2 rows, only one contrast of α's is possible, and to eliminate β's it can only use the data in rows 1 and 2, that are in the same columns. Hence the α-based contrast is

$$\alpha_1 + \tfrac{1}{2}(\gamma_{11} + \gamma_{12}) - \alpha_2 + \tfrac{1}{2}(\gamma_{21} + \gamma_{22})$$

or any (scalar) multiple of this, $f = L_2[\alpha_1 - \alpha_2 + \tfrac{1}{2}(\gamma_{11} + \gamma_{12}) - \tfrac{1}{2}(\gamma_{21} + \gamma_{22}]$, as in the output.

Because there is only one possible contrast of α's the orthogonality argument doesn't get used. But see TYPE III for B, on p. 54.

Using LM 190, Eq. (70), the numerator sum of squares for testing f = 0 is

$$\frac{[\tfrac{1}{2}(\bar{y}_{11\cdot} + \bar{y}_{12\cdot}) - \tfrac{1}{2}(\bar{y}_{21\cdot} + \bar{y}_{22\cdot})]^2}{\tfrac{1}{4}[\tfrac{1}{3} + \tfrac{1}{2} + \tfrac{1}{2} + \tfrac{1}{2}]} = \frac{[4 + 5 - (10 + 9)]^2}{\tfrac{1}{3} + \tfrac{3}{2}} = \frac{100}{\frac{11}{6}} \quad \frac{600}{11} = 54.54 ,$$

as seen on output p. 59.

UNBALANCED DATA, TWO WAY CROSSED CLASSIFICATION
WITH INTERACTION SOME N(I,J) = 0
EXAMPLE ADAPTED FROM BU-417-M BY S R SEARLE

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

TYPE III ESTIMABLE FUNCTIONS FOR: B

EFFECT		COEFFICIENTS
INTERCEPT		0
A	1	0
	2	0
B	1	L4
	2	L5
	3	-L4-L5
A*B	1 1	0.75*L4+0.25*L5
	1 2	0.25*L4+0.75*L5
	1 3	-L4-L5
	2 1	0.25*L4-0.25*L5
	2 2	-0.25*L4+0.25*L5

TYPE III for B in the sequence A. B, AB.

$R^{*}(\beta|\mu,\alpha,\gamma)$ is explained. (SSB_w when all cells filled).

We need β-based contrasts that are

(i) contrasts of β's, and a "mess" of γ's

(ii) involve no α's

(iii) are orthogonal to AB contrasts of $R(\gamma|\mu,\alpha,\beta)$.

From the pattern of filled cells, suitable contrasts satisfying (i) and (ii) are

$$\beta_1 - \beta_2 + \tfrac{1}{2}(\gamma_{11} + \gamma_{21}) - \tfrac{1}{2}(\gamma_{12} + \gamma_{22})$$

$$\beta_1 - \beta_3 + \gamma_{11} - \gamma_{13}$$

$$\beta_2 - \beta_3 + \gamma_{12} - \gamma_{13} .$$

These are not LIN. But any b - 1 = 2 LIN combination of them can be used. Let a general linear combination of them be

$$f = m_1[\beta_1 - \beta_2 + \tfrac{1}{2}(\gamma_{11} + \gamma_{21}) - \tfrac{1}{2}(\gamma_{12} + \gamma_{22})] + m_2[\beta_1 - \beta_3 + \gamma_{11} - \gamma_{13}] + m_3[\beta_2 - \beta_3 + \gamma_{12} - \gamma_{13}] .$$

Now use (iii); the only contrast suitable for $R(\gamma|\mu,\alpha,\beta)$ is

$$\gamma_{11} - \gamma_{12} - \gamma_{21} + \gamma_{22} .$$

Coefficient of in	β_1	β_2	β_3	γ_{11}	γ_{12}	γ_{13}	γ_{21}	γ_{22}
f	$m_1 + m_2$	$-m_1 + m_3$	$-m_2 - m_3$	$\tfrac{1}{2}m_1 + m_2$	$-\tfrac{1}{2}m_1 + m_3$	$-m_2 - m_3$	$\tfrac{1}{2}m_1$	$-\tfrac{1}{2}m_1$
γ-contrast	0	0	0	1	-1	0	-1	1

Orthogonality ⇒ $\tfrac{1}{2}m_1 + m_2 + \tfrac{1}{2}m_1 - m_3 - \tfrac{1}{2}m_1 - \tfrac{1}{2}m_1 = 0$;

i.e., $m_2 = m_3$.

The degrees of freedom for B are 2; so there will be 2 arbitrary values for them m's: choose them so that the β_1, β_2 part of f is $L_4\beta_1 + L_5\beta_2$;

i.e., $m_1 + m_2 = L_4$

$m_1 + m_3 = L_5$.

Hence with $m_2 = m_3$,

$$m_1 = \tfrac{1}{2}(L_4 - L_5) \qquad m_2 = \tfrac{1}{2}(L_4 + L_5) = m_3 .$$

Putting these into F yields the output.

Example

Term in γ_{11} is $\tfrac{1}{2}m_1 + m_2 = \tfrac{1}{4}(L_4 - L_5) + \tfrac{1}{2}(L_4 + L_5) = \tfrac{3}{4}L_4 + \tfrac{1}{4}L_5$.

DEPENDENT VARIABLE: Y

TYPE IV ESTIMABLE FUNCTIONS FOR: A

EFFECT		COEFFICIENTS
INTERCEPT		0
A	1	L2
	2	-L2
B	1	0
	2	0
	3	0
A*B	1 1	0.5*L2
	1 2	0.5*L2
	1 3	0
	2 1	-0.5*L2
	2 2	-0.5*L2

NOTE: OTHER TYPE IV ESTIMABLE FUNCTIONS EXIST.

TYPE IV functions do not have the purpose of explaining some pre-ordained sum of squares, as do TYPES I, II and III. TYPE IV functions are estimable functions that are "contrasts", derived from non-unique, balanced subsets of filled cells of the data. It is their non-uniqueness which gives rise to the NOTE that follows each output. The choice of which balanced subsets can be used is arbitrary, although limited by the pattern of filled cells.

Example

Filled cells

✓	✓	✓
✓	✓	

n_{ij}'s

3	2	1	6
2	2	-	4
5	4	1	10

means

4	5	5
10	9	-

For α-based contrasts, possible balanced subsets of filled cells are

(i) cells 11, 21 (ii) cells 12, 22 (iii) cells 11, 12, 21 and 22.

Obviously, in some general sense, (iii) is the most efficient. The corresponding α-based contrast is $\alpha_1 - \alpha_2 + \frac{1}{2}(\gamma_{11} + \gamma_{12} - \gamma_{21} - \gamma_{22})$, a general form of which is

$$f = L_2[\alpha_1 - \alpha_2 + \tfrac{1}{2}(\gamma_{11} + \gamma_{12} - \gamma_{21} - \gamma_{22})]$$

which is the output. Using (93) on LM 305 the numerator sum of squares for H: $f = 0$ is

$$(\tfrac{1}{2})^2(4 + 5 - 10 - 9)^2/\tfrac{1}{4}(\tfrac{1}{3} + \tfrac{1}{2} + \tfrac{1}{2} + \tfrac{1}{2}) = \frac{100}{11/6} = 54.5455,$$ as shown on output p. 59.

UNBALANCED DATA, TWO WAY CROSSED CLASSIFICATION
WITH INTERACTION SOME N(I,J) = 0
EXAMPLE ADAPTED FROM BU-417-M BY S R SEARLE

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

TYPE IV ESTIMABLE FUNCTIONS FOR: B

EFFECT		COEFFICIENTS
INTERCEPT		0
A	1	0
	2	0
B	1	L4
	2	L5
	3	-L4-L5
A*B	1 1	L4
	1 2	L5
	1 3	-L4-L5
	2 1	0
	2 2	0

NOTE: OTHER TYPE IV ESTIMABLE FUNCTIONS EXIST.

For β-based contrasts, possible balanced subsets of the data are

(i) cells 11, 12 and 11, 13 (ii) cells 21, 22 and 11, 13 or 12, 13 or 11, 12, 13

or (iii) cells 11, 12, 21, 22 and 11, 13 or 12, 13 or 11, 12, 13.

The output is based on subset (i), for which two β-based contrasts are

$$H: \left\{ \begin{matrix} \beta_1 - \beta_3 + \gamma_{11} - \gamma_{13} \\ \beta_1 - \beta_3 + \gamma_{12} - \gamma_{13} \end{matrix} \right\} = 0 \, .$$

From (93) of LM 305 the numerator sum of squares for this is

$$(4 - 5 \quad 5 - 5) \begin{bmatrix} \frac{1}{3} + \frac{1}{1} & \frac{1}{1} \\ \frac{1}{1} & \frac{1}{2} + \frac{1}{1} \end{bmatrix}^{-1} \begin{bmatrix} 4 - 5 \\ 5 - 5 \end{bmatrix} = \frac{(-1)^2(1\frac{1}{2})}{\frac{4}{3}\left(\frac{3}{2}\right) - 1^2} = 1.5$$

as shown on output p. 59.

UNBALANCED DATA, TWO WAY CROSSED CLASSIFICATION
WITH INTERACTION SOME N(I,J) = 0
EXAMPLE ADAPTED FROM BU-417-M BY S R SEARLE

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	4	62.50000000	15.62500000	3.00	0.1294	0.706215	35.0823
ERROR	5	26.00000000	5.20000000		STD DEV		Y MEAN
CORRECTED TOTAL	9	88.50000000			2.28035085		6.50000000

SOURCE	DF		TYPE I SS	F VALUE	PR > F	DF		TYPE II SS	F VALUE	PR > F
A	1	$R(\alpha\|\mu)$	60.00000000	11.54	0.0193	1	$R(\alpha\|\mu,\beta)$	57.01818182	10.97	0.0212
B	2	$R(\beta\|\mu,\alpha)$	0.31818182	0.03	0.9700	2	$R(\beta\|\mu,\alpha)$	0.31818182	0.03	0.9700
A*B	1	$R(\gamma\|\mu,\alpha,\beta)$	2.18181818	0.42	0.5457	1	$R(\gamma\|\mu,\alpha,\beta)$	2.18181818	0.42	0.5457

SOURCE	DF		TYPE III SS	F VALUE	PR > F	DF	TYPE IV SS	F VALUE	PR > F
A	1	$R^*(\alpha\|\mu,\beta,\gamma)_\Sigma$	54.54545455	10.49	0.0230	1*	54.54545455	10.49	0.0230
B	2	$R^*(\beta\|\mu,\alpha,\gamma)_\Sigma$	0.20754717	0.02	0.9803	2*	1.50000000	0.14	0.8692
A*B	1		2.18181818	0.42	0.5457	1	2.18181818	0.42	0.5457

* NOTE: OTHER TYPE IV TESTABLE HYPOTHESES EXIST WHICH MAY YIELD DIFFERENT SS.

See output p. 57 (B, 1.50000000)
See output p. 56 (A, 54.54545455)

PARAMETER		ESTIMATE		T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
INTERCEPT		9.00000000	B	2.79	0.0384	3.22490310
A	1	-4.00000000	B	-1.75	0.1398	2.28035085
	2	0.00000000	B	.	.	.
B	1	1.00000000	B	0.28	0.7926	3.60555128
	2	0.00000000	B	0.00	1.0000	2.79284801
	3	0.00000000	B	.	.	.
A*B	1 1	-2.00000000	B	-0.65	0.5457	3.08760965
	1 2	0.00000000	B	.	.	.
	1 3	0.00000000	B	.	.	.
	2 1	0.00000000	B	.	.	.
	2 2	0.00000000	B	.	.	.

NOTE: AN INFINITE NUMBER OF SOLUTIONS TO THE NORMAL EQUATIONS EXIST. ESTIMATES FOLLOWED BY THE LETTER B ARE BIASED. REFER TO THE GENERAL FORM OF ESTIMABLE FUNCTIONS TO SEE WHAT THE EXPECTED VALUE OF THE BIASED ESTIMATORS ARE.

UNBALANCED DATA, TWO WAY CROSSED CLASSIFICATION
WITH INTERACTION SOME N(I,J) = 0
EXAMPLE ADAPTED FROM BU-417-M BY S R SEARLE

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

TYPE I ESTIMABLE FUNCTIONS FOR: B

EFFECT		COEFFICIENTS
INTERCEPT		0
B	1	L2
	2	L3
	3	-L2-L3
A	1	-0.4*L2-0.5*L3
	2	0.4*L2+0.5*L3
A*B	1 1	0.6*L2
	1 2	0.5*L3
	1 3	-L2-L3
	2 1	0.4*L2
	2 2	0.5*L3

TYPE I for B in the sequence B, A AB.

Explains $R(\beta|\mu)$.

This tests $H: \beta_j + \frac{1}{n_{.j}} \sum_{i=1}^{a} n_{ij}(\alpha_i + \gamma_{ij})$ all equal.

n_{ij}'s			
3	2	1	6
2	2	-	4
5	4	1	10

$$H: \left\{ \begin{array}{l} \beta_1 + \frac{1}{5}(3\alpha_1 + 2\alpha_2 + 3\gamma_{11} + 2\gamma_{21}) \\ \beta_2 + \frac{1}{4}(2\alpha_1 + 2\alpha_2 + 2\gamma_{12} + 2\gamma_{22}) \\ \beta_3 + \frac{1}{1}(\alpha_1 + \gamma_{13}) \end{array} \right\} \text{ equal}$$

A general β-based contrast of these is

$$L_2[\beta_1 + \tfrac{1}{5}(3\alpha_1 + 2\alpha_2 + 3\gamma_{11} + 2\gamma_{21})] + L_3[\beta_2 + \tfrac{1}{4}(2\alpha_1 + 2\alpha_2 + 2\gamma_{12} + 2\gamma_{22})] + (-L_2 - L_3)[\beta_3 + (\alpha_1 + \gamma_{13})],$$

which yields the output; e.g.,

term in α_1 is $\frac{3}{5}L_2 + \frac{2}{4}L_3 + (-L_2 - L_3) = -0.4L_2 - 0.5L_3$.

term in γ_{11} is $\frac{3}{5}L_2 = 0.6L_2$.

UNBALANCED DATA, TWO WAY CROSSED CLASSIFICATION
WITH INTERACTION SOME N(I,J) = 0
EXAMPLE ADAPTED FROM BU-417-M BY S R SEARLE

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE	DF		SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	4		62.50000000	15.62500000	3.00	0.1294	0.706215	35.0823
ERROR	5	SSE	26.00000000	5.20000000		STD DEV		Y MEAN
CORRECTED TOTAL	9		88.50000000			2.28035085		6.50000000

SOURCE	DF		TYPE I SS	F VALUE	PR > F
B	2	$R(\beta \mid \mu)$	3.30000000	0.32	0.7418
A	1	$R(\alpha \mid \mu,\beta)$	57.01818182	10.97	0.0212
A*B	1	$R(\gamma \mid \mu,\alpha,\beta)$	2.18181818	0.42	0.5457