ANNOTATED COMPUTER OUTPUT FOR ANALYSES OF UNBALANCED DATA: GENSTAT ANOVA and REGRESSION

BU-644-M by Revised May, 1979

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 paper reports the running of those data sets on the ANOVA and REGRESSION routines in GENSTAT 3.09. 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 later data sets 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-644-M in the Biometrics Unit Mimeo Series.

* Supported by a New Zealand National Research Advisory Council Fellowship.

Data Set (1): Balanced Data

2-way crossed classification, 3 rows, 4 columns, and
2 observations per cell

					Total	(n)	Mean
	10	16	12	9			
	14	22	18	19			
	24	38	30	28	120	(8)	15
	23	17	24	18			
	25	21	32	24			
	48	38	56	42	184	(8)	23
	13	8	16	7			
	17	12	12	19			
	30	20	28	26	104	(8)	13
Total	102	96	114	96	408		
(n)	(6)	(6)	(6)	(6)		(24)	
Mean	17	16	19	16			17

Sums of Squares

$R(\mu) = 6936$

$R(\mu,\alpha) = 7384$

$R(\mu,\beta) = 6972$

$R(\mu,\alpha,\beta,\gamma) = 7556$

$\underset{\sim}{y}'\underset{\sim}{y} = 7806$

Analysis of Variance

Term	d.f.	S.S.
Mean	1	6936
Rows	2	448
Columns	3	36
Interaction	6	136
Error	12	250
Total	24	7806

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

$R(\mu) = 1296$

$R(\mu,\alpha) = 1458$

$R(\mu,\beta) = 1444\frac{1}{2}$

$R(\mu,\alpha,\beta) = 1716$

$\underset{\sim}{y}'\underset{\sim}{y} = 1728$

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-668-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: $\underset{\sim}{b}^{o'} = [0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 8 \quad 6 \quad 2 \quad 8 \quad 6 \quad 12]$

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

2-way crossed classification, 2 rows and 3 columns

				Totals
	2	4	5	
	4	6		
	6			
	12(3)4	10(2)5	5(1)5	27(6)4$\frac{1}{2}$
	12	11	-	
	8	7		
	20(2)10	18(2)9	-	38(4)9$\frac{1}{2}$
Totals:	32(5)6.4	28(4)7	5(1)5	65(10)6$\frac{1}{2}$

Sums of Squares

$R(\mu)$	$= 422\frac{1}{2}$
$R(\mu,\alpha)$	$= 482\frac{1}{2}$
$R(\mu,\beta)$	$= 425.8$
$R(\mu,\alpha,\beta)$	$= 482.818$
$R(\mu,\alpha,\beta,\gamma)$	$= 485$
$\underset{\sim}{y}'\underset{\sim}{y}$	$= 511$

Analyses of Variance

a. Rows before columns			b. Columns before rows		
Term	d.f.	S.S.	Term	d.f.	S.S.
$R(\mu)$	1	422.5	$R(\mu)$	1	422.5
$R(\alpha\|\mu)$	1	60	$R(\beta\|\mu)$	2	3.3
$R(\beta\|\mu,\alpha)$	2	.318	$R(\alpha\|\mu,\beta)$	1	57.018
$R(\gamma\|\mu,\alpha,\beta)$	1	2.182	$R(\gamma\|\mu,\alpha,\beta)$	1	2.182
SSE	5	26	SSE	5	26
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 (5): Unbalanced Data, empty cells – a bigger case

2-way crossed classification, 3 rows and 4 columns

[LM, pp. 287, 294]

					Totals
	8 13 9	-	12	7 11	
	30(3)10	-	12(1)12	18(2)9	60(6)10
	6 12	12 14	-	-	
	18(2)9	26(2)13	-	-	44(4)11
	-	9 7	14 16	10 14 11 13	
	-	16(2)8	30(2)15	48(4)12	94(8)$11\frac{3}{4}$
Totals:	48(5)9.6	42(4)$10\frac{1}{2}$	42(3)14	66(6)11	198(18)11

Analyses of Variance

Sums of Squares		a. Rows before columns: Term	d.f.	S.S.	b. Columns before rows: Term	d.f.	S.S.
$R(\mu)$	$= 2178$	$R(\mu)$	1	2178	$R(\mu)$	1	2178
$R(\mu,\alpha)$	$= 2188\frac{1}{2}$	$R(\alpha\|\mu)$	2	$10\frac{35}{70}$	$R(\beta\|\mu)$	3	$37\frac{56}{70}$
$R(\mu,\beta)$	$= 2215.8$	$R(\beta\|\mu,\alpha)$	3	$36\frac{55}{70}$	$R(\alpha\|\mu,\beta)$	2	$9\frac{34}{70}$
$R(\mu,\alpha,\beta)$	$= 2225\frac{2}{7}$	$R(\gamma\|\mu,\alpha,\beta)$	2	$34\frac{50}{70}$	$R(\gamma\|\mu,\alpha,\beta)$	2	$34\frac{50}{70}$
$R(\mu,\alpha,\beta,\gamma)$	$= 2260$	SSE	10	56	SSE	10	56
$\underset{\sim}{y}'\underset{\sim}{y}$	$= 2316$	SST	18	2316	SST	18	2316

Parameter vector: $\underset{\sim}{b}' = [\mu \quad \underset{\sim}{\alpha}'_{1x3} \quad \underset{\sim}{\beta}'_{1x4} \quad \gamma_{11} \quad \gamma_{13} \quad \gamma_{14} \quad \gamma_{21} \quad \gamma_{22} \quad \gamma_{32} \quad \gamma_{33} \quad \gamma_{34}]$

Solution vector: $\underset{\sim}{b}^{o'} = [0 \quad \underset{\sim}{0} \quad \underset{\sim}{0} \quad 10 \quad 12 \quad 9 \quad 9 \quad 13 \quad 8 \quad 15 \quad 12]$

Data Set (6): Covariance, 1-way classification

3 groups, 1 covariate, unbalanced data

[LM, pp. 353-359]

	Group 1		Group 2		Group 3	
	Variable y	Covariate z	Variable y	Covariate z	Variable y	Covariate z
	74	3	76	2	85	4
	68	4	80	4	93	6
	77	2				
Totals:	219	9	156	6	178	10
Means:	73	3	78	3	89	5

Sums of Squares and Products

		yy	zz	yz
$R(\mu) \equiv$ SSM		43,687	$89\frac{2}{7}$	1975
$R(\mu,\alpha)$ = SSR		43,997	95	2015
Total = SST		44,079	101	2018

	d.f.			
$R(\mu)$	1	43,687	$89\frac{2}{7}$	1975
$R(\alpha\|\mu)$	2	310	$5\frac{5}{7}$	40
SSE	4	82	6	3
SST	7	44,079	101	2018

(1) Ignoring covariate: Parameter vector is $\underset{\sim}{a}^{o\prime} = [\mu \quad \alpha_1 \quad \alpha_2 \quad \alpha_3]$

Solution vector is $\underset{\sim}{a}^{*\prime} = [0 \quad 73 \quad 78 \quad 89]$

(2) Pooled within-group regression: $y_{ij} = \mu + \alpha_i + bz_{ij} + e_{ij}$ [(43), LM, p. 353]

Estimated slope (pooled): $\hat{b} = \frac{1}{2}$ [(42), LM, p. 353]

Solution vector for groups: $\underset{\sim}{a}^{o\prime} = [0 \quad 71\frac{1}{2} \quad 76\frac{1}{2} \quad 86\frac{1}{2}]$ [(44), LM, p. 353]

Analyses of Variance

a. Fitting groups before covariate. [Table 8.6a, LM, p. 354]

Source of Variation	d.f.	Sum of Squares
Mean	1	$R(\mu) = 43,687$
α-classes (after mean)	2	$R(\alpha\|\mu) = 310$
Covariate (pooled within-class regression)	1	$R(b\|\mu,\alpha) = 1.5$
Residual error	3	$SSE = 80.5$
Total	7	$SST = 44,079$

b. Fitting covariate before groups. [Table 8.6b, LM, p. 355]

Source of Variation	d.f.	Sum of Squares
Mean	1	$R(\mu) = 43,687$
Covariate (after mean)	1	$R(b\|\mu) = 157.8$
α-classes (after mean and covariate)	2	$R(\alpha\|\mu,b) = 153.7$
Residual error	3	$SSE = 80.5$
Total	7	$SST = 44,079$

(3) Intra-class regression: $y_{ij} = \mu + \alpha_i + b_i z_{ij} + e_{ij}$

Estimated slopes: $\hat{b}_1 = -4\frac{1}{2}$ $\hat{b}_2 = 2$ $\hat{b}_3 = 4$ [LM, p. 358]

Solution vector for groups:

$$\underset{\sim}{a}^{o} = \begin{bmatrix} 0 \\ 73 - (-4\frac{1}{2})(3) \\ 78 - 2(3) \\ 89 - 4(5) \end{bmatrix} = \begin{bmatrix} 0 \\ 86\frac{1}{2} \\ 72 \\ 69 \end{bmatrix}$$ [See (52), LM, p. 357]

Analysis of Variance [LM, p. 359]

Source of Variation	d.f.		Sum of Squares	
Mean	1		$R(\mu) = 43,687$	
α-classes (after mean)	2		$R(\alpha\|\mu) = 310$	
Covariate (within-class)	3		$R(b\|\mu,\alpha) = 80.5$	
Pooled		1		$R(b\|\mu,\alpha) = 1.5$
Difference		2		Difference $= 79$
Residual error	1		$SSE = 1.5$	
Total	7		$SST = 44,079$	

Data Set (7): Covariance, 2-way crossed classification

3 rows, 4 columns, unbalanced data with empty cells
[LM, Exercise 8.12; Solutions Manual, p. 148-150]

y	z	y	z	y	z	y	z
8	2			12	7	7	3
13	4	-				11	5
9	3						
6	5	12	6				
12	3	14	4	-		-	
		9	6	14	6	10	4
-		7	2	16	8	14	6
						11	5
						13	7

(1) Ignoring covariate: Data and analyses are those of Data Set (5).

(2) Different slope for each row: $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + b_i z_{ijk} + e_{ijk}$

Estimated slopes: $\hat{b}_1 = 2\frac{1}{4}$ $\quad \hat{b}_2 = -2$ $\quad \hat{b}_3 = 4/5$

Parameter vector: $\underset{\sim}{\mu}' = \{\mu_{ij}\} = \{\mu + \alpha_i + \beta_j + \gamma_{ij}\}$

Solution vector:

No covariate: $\underset{\sim}{\mu}^{*'} = \{\bar{y}_{ij.}\} = [10 \quad 12 \quad 9 \quad 9 \quad 13 \quad 8 \quad 15 \quad 12]$

Covariate: $\underset{\sim}{\mu}^{o'} = \{\bar{y}_{ij.} - b_i \bar{z}_{ij.}\}$

$\{\bar{z}_{ij.}\} = [\,3 \quad 7 \quad 4 \quad 4 \quad 5 \quad 4 \quad 7 \quad 5\frac{1}{2}]$

$\{b_i \bar{z}_{ij.}\} = [6\frac{3}{4} \quad 15\frac{3}{4} \quad 9 \quad -8 \quad -10 \quad 3\frac{1}{5} \quad 5\frac{3}{5} \quad 4\frac{2}{5}]$

$\underset{\sim}{\mu}^{o'} = [3\frac{1}{4} \quad -3\frac{3}{4} \quad 0 \quad 17 \quad 23 \quad 4\frac{4}{5} \quad 9\frac{2}{5} \quad 7\frac{3}{5}]$

Note: Errors in Solutions Manual:

Page 149, last entry on bottom line: $\frac{13}{15}$ should be $\frac{12}{15} = \frac{4}{5}$.

Page 150: $R(\underset{\sim}{b}|a) = [2\tfrac{1}{4} \quad -2 \quad \tfrac{4}{5}]\begin{bmatrix} 9 \\ -8 \\ 12 \end{bmatrix} = 45\frac{17}{20}$ (two places)

$SSE = 10\frac{3}{20}$

Page 151: $R(\underset{\sim}{b}|\mu) = 53.97$

$R(\alpha|\mu,\underset{\sim}{b}) = 73.88$

$SSE = 10.15$

Check on SSE:

$$\underset{\sim}{b}^{o'}\underset{\sim}{X}'\underset{\sim}{y} = 3\tfrac{1}{4}(30) - 3\tfrac{3}{4}(12) + 0(18) + 17(18) + 23(26) + 4\tfrac{4}{5}(16) + 9\tfrac{2}{5}(30) + 7\tfrac{3}{5}(48)$$

$$+ 2\tfrac{1}{4}[8(2) + 13(4) + 9(3) + 12(7) + 7(3) + 11(5)]$$

$$- 2[6(5) + 12(3) + 12(6) + 14(4)]$$

$$+ \tfrac{4}{5}[9(6) + 7(2) + 14(6) + 16(8) + 10(4) + 14(6) + 11(5) + 13(7)]$$

$$= 1725\frac{1}{10} - 45 + 573\frac{3}{4} - 388 + 440$$

$$= 2305\frac{17}{20} .$$

$SSE = SST - 2305\frac{17}{20} = 2316 - 2305\frac{17}{20} = 10\frac{3}{20}$, as expected.

Analyses of Variance [Solutions Manual, pp. 151-152]

a. Fitting factors before covariates.

Source of Variation	d.f.		Sum of Squares	
Factors and mean	8		2260	
Mean		1		2178
Factors after mean		7		82
Residual	10		56	
Covariates after factors		3		$45\frac{17}{20}$
Error		7		$10\frac{3}{20}$
Total	18		2316	

b. Fitting covariates before factors.

Source of Variation	d.f.	Sum of Squares
Mean	1	2178
Covariates after mean	3	53.97
Factors after mean and covariates	7	73.88
Error	7	10.15
Total	18	2316

References

LM: Searle, S. R. [1971]. Linear Models, Wiley, New York.

BU-343: Searle, S. R. [1979]. Arbitrary hypotheses in linear models with unbalanced data. Communications in Statistics, A8, No. 8. (in press).

BU-668-M: Searle, S. R., Speed, F. M., and Henderson, H. V. [1979]. Some computational and model equivalences in analyses of variance of unequal-subclass-numbers data. Paper No. BU-668-M, Biometrics Unit, Cornell University.

BU-672-M: Searle, S. R., Milliken, G. A., and Speed, F. M. [1979]. Expected marginal means in the linear model. Paper No. BU-672-M, Biometrics Unit, Cornell University.

```
GENSTAT V     MARK 3.09
(C)1976 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STATION)
```

GENSTAT ANOVA - Data Set 1

```
   1  'REFERENCE' SET_1
   2  'CAPTION'
  -3            ''
  -4              DATA SET 1
  -5              BALANCED 3*4 FACTORIAL TREATMENT DESIGN
  -6              WITH 2 OBSERVATIONS PER CELL IN CRD DESIGN
   7            ''
   8  'UNITS'     $ 24
   9  'FACTORS'   A $ 3
  10    :         B $ 4
  11  'GENERATE'  A,B,2
  12  'READ/P'    Y
  13  'PRINT/P'   A,B,Y   $ 10.0
  14  'TREATMENTS' A*B
  15  'ANOVA'     Y
  16  'RUN'
```

GENSTAT program prepared by the user.

```
DATA SET 1
BALANCED 3*4 FACTORIAL TREATMENT DESIGN
WITH 2 OBSERVATIONS PER CELL IN CRD DESIGN
```

```
  Y  0  MNMINMAX      17.0000      7.0000      32.0000     24  VALUES     0 MISSING
```

Variable Mean — Mean — Minimum — Maximum — No. of observations — No. of "missing" observations.

A	B	Y
1	1	10
1	1	14
1	2	16
1	2	22
1	3	12
1	3	18
1	4	9
1	4	19
2	1	23
2	1	25
2	2	17
2	2	21
2	3	24
2	3	32
2	4	18
2	4	24
3	1	13
3	1	17
3	2	8
3	2	12
3	3	16
3	3	12
3	4	7
3	4	19

***** ANALYSIS OF VARIANCE *****

VARIATE: Y

SOURCE OF VARIATION		DF	SS	SS%	MS	VR = Variance Ratio
UNITS STRATUM						
A	SSA_m	2	448.00	51.49	224.00	10.752
B	SSB_m	3	36.00	4.14	12.00	0.576
A.B	$SSAB_m$	6	136.00	15.63	22.67	1.088
RESIDUAL	SSE	12	250.00	28.74	20.83	
TOTAL	SST_m	23	870.00	100.00	37.83	
GRAND TOTAL	(N-1)	23	870.00	100.00		

Mean squares divided by mean square for RESIDUAL; e.g., 22.67/20.83 = 1.088. These may, or may not, be appropriate as F-statistics, depending on the underlying model which is specified by the user.

GRAND MEAN 17.00
TOTAL NUMBER OF OBSERVATIONS 24

Sum of squares corrected for the mean: $SST_m = SST - SSM$. But $SSM = N\bar{y}^2$ can be calculated; e.g.,
$SSM = N\bar{y}^2 = 24(17^2) = 6936$
$SST = SST_m + SSM = 870 + 6936 = 7806$.

This differs from TOTAL in analyses of nested designs.

***** TABLES OF MEANS *****

VARIATE: Y

GRAND MEAN 17.00

A	1	2	3
	15.00	23.00	13.00

B	1	2	3	4
	17.00	16.00	19.00	16.00

B	1	2	3	4
A				
1	12.00	19.00	15.00	14.00
2	24.00	19.00	28.00	21.00
3	15.00	10.00	14.00	13.00

***** STANDARD ERRORS OF DIFFERENCES OF MEANS *****

TABLE	A	B	A B
REP	8	6	2
SED	2.282	2.635	4.564

No. of observations in the means of each level of each factor.

Standard error of a difference between means of any two levels of factor. As examples,

$$2.282 = \sqrt{\frac{2(20.8)}{8}} \quad \text{and} \quad 2.635 = \sqrt{\frac{2(20.8)}{6}}$$

***** STRATUM STANDARD ERRORS AND COEFFICIENTS OF VARIATION *****

STRATUM	DF	SE	CV%
UNITS	12	4.564	26.8

SE = Standard error = $\sqrt{\hat{\sigma}_e^2} = \sqrt{20.8} = 4.564$

CV% = Coefficient of variation = $(100SE/\bar{y})\% = (456.4/17)\% = 26.8\%$.

21 *CLOSE*

```
GENSTAT V     MARK 3.09
(C)1976 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STATION)

'REFERENCE' SET_2
'CAPTION'
              ''
                DATA SET 2
                UNBALANCED DATA, NO INTERACTION, N(I,J) = 0 OR 1
                LINEAR MODELS BY S R SEARLE  PAGE 262
              ''
'UNITS'         $ 12
'READ/P'        Y
'NAME'          LEVELSA = X,Y,Z,W
  :             LEVELSB = A,B,C
'FACTOR'        A $ LEVELSA
  :             B $ LEVELSB
'GENERATE'      A,B
'PRINT/P'       A,B,Y $10.0
'TREATMENTS'       A + B
'ANOVA'         Y
'RUN'


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

  Y  0  MNMINMAX        12.0000        3.0000        24.0000     12  VALUES     3 MISSING
```

A	B	Y
X	A	18
X	B	12
X	C	24
Y	A	* ←
Y	B	* ←
Y	C	9
Z	A	3
Z	B	* ←
Z	C	15
W	A	6
W	B	3
W	C	18

The program is designed for balanced data. It treats unbalanced data as balanced data that has missing observations. The user has to indicate those missing observations in the data cards (using blank, ampersand, minus, M, or asterisk), sufficient in number to make the data balanced. The program then uses an iterative technique to estimate those "missing" values, puts the estimated missing values (EMV's) into the data and analyzes the now-balanced data, correcting the degrees of freedom for the number of "missing" values. This is a 2-stage procedure: first, it derives estimated missing values (EMV's) and second, it analyzes the now-balanced data with the EMV's included. Properties of this procedure are discussed, for example, by Cochran and Cox [1957, Sec. 3.7].

Data with EMV's included

EMV's				
	18	12	24	18
	-.95	-4.96	9	1.03
	3	0.02	15	6.01
	6	3	18	9
	6.51	2.52	$16\frac{1}{2}$	8.51

***** ANALYSIS OF VARIANCE *****

VARIATE: Y

SOURCE OF VARIATION		DF(MV)	SS	SS%	MS	VR
UNITS STRATUM						
A		3	457.498	105.90	152.499	38.121
B		2	414.996	96.06	207.498	51.870
RESIDUAL	SSE	3(3)	12.001	2.78	4.000	
TOTAL		8	884.495	204.74	110.562	
GRAND TOTAL		8	884.495	204.74		

Number of "missing" values

This is an approximate analysis. It is the Anova of the data with EMV's included, and degrees of freedom adjusted accordingly.

This is not 100% because $\Sigma y_{i}^{2} - N\bar{y}^{2}$ of just the actual data (not including EMV's) has been used as the denominator. For example.

$$\frac{884.495}{1728 - 1296} = \frac{884.495}{432} = 204.74\% .$$

ESTIMATED GRAND MEAN	8.51
TOTAL NUMBER OF OBSERVATIONS	12
NUMBER OF MISSING VALUES	3
MAXIMUM NUMBER OF ITERATIONS	10

This is the only sum of squares that agrees with the correct analysis. Others always exceed the correct ones.

UNIT NUMBER	ESTIMATED VALUE
4	-0.95
5	-4.96
8	0.02

Iterated estimated missing values. Apart from rounding error (they should be -1, -5 and 0), they are the same as the usual estimated missing values derived from minimizing SSE (LM p. 362). For iterative technique see Payne and Wilkinson, Applied Statistics, 26, 1977.

> Note: The minimizing procedure used to obtain estimated missing values (EMV's) $\mathbf{m}$, is to minimize the SSE for both the data and the EMV's:
>
> $$(\mathbf{y} - \mathbf{Xb})'(\mathbf{y} - \mathbf{Xb}) + (\mathbf{m} - \mathbf{Zb})'(\mathbf{m} - \mathbf{Zb}).$$
>
> Minimizing gives:
>
> w.r.t. $\mathbf{b}$: $\quad \mathbf{X'Xb}^{\circ} + \mathbf{Z'Zb}^{\circ} = \mathbf{X'y} + \mathbf{Z'y}$;
>
> w.r.t. $\mathbf{m}$: $\quad \mathbf{m} = \mathbf{Zb}^{\circ}$.
>
> Therefore $\quad \mathbf{Z'm} = \mathbf{Z'Zb}^{\circ}$
>
> and so $\quad \mathbf{X'Xb}^{\circ} = \mathbf{X'y}$.
>
> Hence $\mathbf{b}^{\circ}$ is a solution of the normal equations for $E(\mathbf{y}) = \mathbf{Xb}$.

***** TABLES OF MEANS *****

These are estimated expected values; e.g., $\widehat{E(\bar{y}_{ij})} = \frac{1}{a}\sum_{i}(\hat{\mu} + \hat{\alpha}_{i} + \hat{\beta}_{j})$, which for balanced data are simply sample means.

VARIATE: Y

GRAND MEAN 8.51

This is the estimated grand mean as given 11 lines earlier.

A	X	Y	Z	W
	18.00	1.03	6.01	9.00

B	A	B	C
	6.51	2.52	16.50

In this example, means shown here for X, W, and C are the same as in the data, because there are no empty cells in those classes — and hence no EMV's. Other means shown here include those EMV's. For example, for Y: $(-.95 - 4.96 + 9)/3 = 1.03$.

***** STANDARD ERRORS OF DIFFERENCES OF MEANS *****

TABLE	A	B
REP	3	4
SED	1.633	1.414

(NOT ADJUSTED FOR MISSING VALUES)

These results also assume the "missing" values are not missing; e.g. that there are 3 observations for each row; and hence also, $1.633 = \sqrt{2(4)/3} = 1.63$. Hence, the standard errors are underestimated for rows (columns) with missing values. A warning is given in the output indicating this.

***** STRATUM STANDARD ERRORS AND COEFFICIENTS OF VARIATION *****

STRATUM	DF	SE	CV%
UNITS	3	2.000	23.5

Note: The solution vector for $\mathbf{b}$ based on (i) the "balanced" data including EMV's (rounded to their correct values, -1, -5, 0) and (ii) the "usual" constraints, is

$$\mathbf{b}^{*\prime} = [8\tfrac{1}{2} \,\vdots\, 9\tfrac{1}{2} \;\; -7\tfrac{1}{2} \;\; -2\tfrac{1}{2} \;\; \tfrac{1}{2} \,\vdots\, -2 \;\; -6 \;\; 8].$$

This is also a solution vector for the unbalanced data.

```
GENSTAT V    MARK 3.09
(C)1976 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STATION)
```

GENSTAT ANOVA - Data Set 3

```
'CAPTION'
        ''
              DATA SET 3
              UNBALANCED DATA, TWO WAY CROSSED CLASSIFICATION
                 WITH INTERACTION  N(I,J) > 0
                     EXAMPLE FROM BU-608-M BY S R SEARLE
        ''
'UNITS'      $ 12
'READ/P'     Y
'FACTOR'     A $ 2
   :         B $ 3
'GENERATE'   A,B,2
'PRINT/P'    A,B,Y  $10.0
'TREATMENTS' A*B
'ANOVA'      Y
'RUN'

      Y  0  MNMINMAX        7.0000        2.0000        12.0000     12  VALUES     4 MISSING
```

A	B	Y
1	1	7
1	1	9
1	2	6
1	2	*
1	3	2
1	3	*
2	1	8
2	1	*
2	2	4
2	2	8
2	3	12
2	3	*

***** ANALYSIS OF VARIANCE *****

VARIATE: Y

This ANOVA includes the EMV's in the data.

SOURCE OF VARIATION		DF(MV)	SS	SS%	MS	VR
UNITS STRATUM						
A		1	33.328	50.50	33.328	6.666
B		2	7.999	12.12	4.000	0.800
A.B		2	66.659	101.00	33.330	6.666
RESIDUAL	SSE	2(4)	10.000	15.15	5.000	
TOTAL		7	117.967	178.77	16.855	
GRAND TOTAL		7	117.987	178.77		

Data including EMV's in parentheses.

	7	6	2	
	9	(6)	(2)	
Means	8	6	2	$5\frac{1}{3}$
	8	4	12	
	(8)	8	(12)	
Means	8	6	12	$8\frac{2}{3}$
Means	8	6	-	7

```
ESTIMATED GRAND MEAN              7.00
TOTAL NUMBER OF OBSERVATIONS        12
NUMBER OF MISSING VALUES             4
MAXIMUM NUMBER OF ITERATIONS         3
```

UNIT NUMBER	ESTIMATED VALUE
4	6.00
6	2.00
8	8.00
12	12.00

The EMV's are the cell means — because each cell has some data and because the model contains interactions: $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$.

***** TABLES OF MEANS *****

VARIATE: Y

GRAND MEAN 7.00

Because EMV's are the $\bar{y}_{ij.}$'s, this table of means is simply a table of means of cell means.

A	1	2
	5.33	8.67

Example: $5.33 = \frac{1}{3}(8 + 6 + 2) = \sum_{j=1}^{3} \bar{y}_{ij.}/3$.

B	1	2	3
	8.00	6.00	7.00

The same kind of calculation is made here, $\sum_{i=1}^{a} \bar{y}_{ij.}/a$. and in this case, simply by numerical coincidence, these means are identical to the column means $\bar{y}_{.j.}$.

B / A	1	2	3
1	8.00	6.00	2.00
2	8.00	6.00	12.00

These <u>are</u> the cell means; estimating missing values by cell means leaves the cell means unchanged.

***** STANDARD ERRORS OF DIFFERENCES OF MEANS *****

TABLE	A	B	A B
REP	6	4	2
SED	1.291	1.581	2.236

(NOT ADJUSTED FOR MISSING VALUES)

The solution vector based on (i) the "balanced" data including EMV's and (ii) the "usual" constraints; e.g., $\Sigma\alpha_i = 0$, is

$$b^{*\prime} = [7 \;\vdots\; -1\tfrac{2}{3} \quad 1\tfrac{2}{3} \;\vdots\; 1 \quad -1 \quad 0 \;\vdots\; 1\tfrac{2}{3} \quad 1\tfrac{2}{3} \quad -3\tfrac{1}{3} \quad -1\tfrac{2}{3} \quad -1\tfrac{2}{3} \quad 3\tfrac{1}{3}] .$$

This is also a solution vector for the unbalanced data.

***** STRATUM STANDARD ERRORS AND COEFFICIENTS OF VARIATION *****

STRATUM	DF	SE	CV%
UNITS	2	2.236	31.9

```
  60  'REFERENCE' SET_4
 ******* END OF      CAPTION    AT LINE   56    USED 428    LEFT 7572
```

GENSTAT V   MARK 3.09


```
'CAPTION'
           ''
                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
           ''
'UNITS'         $ 18
'READ/P'        Y
'FACTOR'        A $2
   :            B $ 3
'GENERATE'      A,B,3
'PRINT/P'       A,B,Y $10.0
'TREATMENTS'    A*B
'ANOVA'         Y
'RUN'
```

```
   Y  0  MNMINMAX        6.5000        2.0000       12.0000     18  VALUES     8 MISSING
```

A	B	Y
1	1	2
1	1	4
1	1	6
1	2	4
1	2	6
1	2	*
1	3	5
1	3	*
1	3	*
2	1	12
2	1	8
2	1	*
2	2	11
2	2	7
2	2	*
2	3	*
2	3	*
2	3	*

***** ANALYSIS OF VARIANCE *****

VARIATE: Y

SOURCE OF VARIATION		DF(MV)	SS	SS%	MS	VR
UNITS STRATUM						
A		1	113.000	127.68	113.000	21.731
B		2	1.133	1.28	0.567	0.109
A.B		1(1)	3.001	3.39	3.001	0.577
RESIDUAL	SSE	5(7)	26.000	29.38	5.200	
TOTAL		9	143.134	161.73	15.904	
GRAND TOTAL		9	143.134	161.73		

ESTIMATED GRAND MEAN	7.18
TOTAL NUMBER OF OBSERVATIONS	18
NUMBER OF MISSING VALUES	8
MAXIMUM NUMBER OF ITERATIONS	6

UNIT NUMBER	ESTIMATED VALUE
6	5.00
8	5.02
9	5.02
12	10.00
15	9.00
16	10.05
17	10.05
18	10.05

Estimating 2 missing values for row 1, column 3 where there is only one observation.

Estimating an empty cell; i.e., estimating a cell mean for a cell that has no observations for which, because the interaction model is being used, it is not usual to estimate the cell mean. In fact, such cell means are not estimable. (The program, nevertheless, having "estimated" such a mean, uses it enough times as observations, to make the data balanced.) See Rubin, Applied Statistics 21, 136, 1976, especially pages 140-141.

***** TABLES OF MEANS *****

VARIATE: Y

GRAND MEAN 7.18

A	1	2
	4.67	9.68

B	1	2	3
	7.00	7.00	7.53

B	1	2	3
A			
1	4.00	5.00	5.02
2	10.00	9.00	10.05

```
***** STANDARD ERRORS OF DIFFERENCES OF MEANS *****

TABLE                        A            B            A
                                                       B
---------------------------------------------------------
REP                          9            6            3
SED                      1.075        1.317        1.862

(NOT ADJUSTED FOR MISSING VALUES)


***** STRATUM STANDARD ERRORS AND COEFFICIENTS OF VARIATION *****

STRATUM                        DF             SE          CV%

*UNITS*                         5          2.280         31.8



  79  'REFERENCE' SET_5
 ******* END OF        CAPTION     AT LINE    75     USED 442     LEFT 7558
```

```
GENSTAT V    MARK 3.09
(C)1976 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STATION)
```

GENSTAT ANOVA - Data Set 5

Same characteristics as Data Set 4, but more extensive.

```
'CAPTION'
        ''
           DATA SET 5
           UNBALANCED DATA, TWO WAY CROSSED CLASSIFICATION
              WITH INTERACTION  SOME N(I,J) = 0
                LINEAR MODELS BY SRSEARLE PAGE 287
        ''
'UNITS'      $ 48
'READ/P'     Y
'FACTOR'     A $ 3
  :          B $ 4
'GENERATE'   A,B,4
'PRINT/P'    A,B,Y    $ 10.0
'TREATMENTS' A*B
'ANOVA'      Y
'RUN'

     Y  0  MNMINMAX       11.0000       6.0000      16.0000    48  VALUES   30 MISSING
```

A	B	Y
1	1	8
1	1	13
1	1	9
1	1	*
1	2	*
1	2	*
1	2	*
1	2	*
1	3	12
1	3	*
1	3	*
1	3	*
1	4	7
1	4	11
1	4	*
1	4	*
2	1	6
2	1	12
2	1	*
2	1	*
2	2	12
2	2	14
2	2	*
2	2	*
2	3	*
2	3	*
2	3	*
2	3	*
2	4	*
2	4	*
2	4	*
2	4	*
3	1	*
3	1	*
3	1	*
3	1	*
3	2	9
3	2	7
3	2	*
3	2	*
3	3	14
3	3	16
3	3	*
3	3	*
3	4	10
3	4	14
3	4	11
3	4	13

***** ANALYSIS OF VARIANCE *****

VARIATE: Y

SOURCE OF VARIATION		DF(MV)	SS	SS%	MS	VR
UNITS STRATUM						
A		2	52.488	38.03	26.244	4.686
B		3	158.320	114.72	52.773	9.423
A.B		2(4)	65.101	47.17	32.551	5.812
RESIDUAL	SSE	10(26)	56.002	40.58	5.600	
TOTAL		17	331.911	240.52	19.524	
GRAND TOTAL		17	331.911	240.52		

ESTIMATED GRAND MEAN 11.15
TOTAL NUMBER OF OBSERVATIONS 48
NUMBER OF MISSING VALUES 30
MAXIMUM NUMBER OF ITERATIONS 7

UNIT NUMBER	ESTIMATED VALUE
4	10.00
5	8.62
6	8.62
7	8.62
8	8.62
10	11.94
11	11.94
12	11.94
15	9.00
16	9.00
19	9.00
20	9.00
23	13.00
24	13.00
25	15.39
26	15.39
27	15.39
28	15.39
29	12.42
30	12.42
31	12.42
32	12.42
33	9.46
34	9.46
35	9.46
36	9.46
39	8.00
40	8.00
43	15.00
44	15.00

***** TABLES OF MEANS *****

VARIATE: Y

GRAND MEAN 11.15

A	1	2	3
	9.89	12.45	11.12

B	1	2	3	4
	9.49	9.87	14.12	11.14

B A	1	2	3	4
1	10.00	8.62	11.96	9.00
2	9.00	13.00	15.39	12.42
3	9.46	8.00	15.00	12.00

***** STANDARD ERRORS OF DIFFERENCES OF MEANS *****

TABLE	A	B	A B
REP	16	12	4
SED	0.837	0.966	1.673

(NOT ADJUSTED FOR MISSING VALUES)

***** STRATUM STANDARD ERRORS AND COEFFICIENTS OF VARIATION *****

STRATUM	DF	SE	CV%
UNITS	10	2.366	21.2

```
 100  'REFERENCE' SET_6
 ******* END OF       CAPTION    AT LINE   94    USED 487    LEFT 7513
```

GENSTAT V MARK 3.09
(C)1976 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STATION)


```
 103  'REFERENCE' SET_6
 104  'CAPTION'
-105            ''
-106                 DATA SET 6
-107                 UNBALANCED DATA, 1 WAY CLASSIFICATION
-108                 WITH 1 COVARIATE
-109                 LINEAR MODELS BY S R SEARLE PAGE 353
 110            ''
 111  'UNITS'       $ 9
 112  'FACTOR'      A $ 3
 113  'READ/P'      A,Y,X
 114  'PRINT/P'     X,A,Y $ 10.0
 115  'TREATMENTS'    A
 116  'COVARIATES'    X
 117  'ANOVA'       Y
 118  'RUN'
```

```
        DATA SET 6
        UNBALANCED DATA, 1 WAY CLASSIFICATION
        WITH 1 COVARIATE
        LINEAR MODELS BY S R SEARLE PAGE 353

   Y  0  MNMINMAX      79.0000     68.0000     93.0000    9  VALUES    2 MISSING
   X  0  MNMINMAX       3.5714      2.0000      6.0000    9  VALUES    2 MISSING
```

X	A	Y
3	1	74
4	1	68
2	1	77
2	2	76
4	2	80
*	2	*
4	3	85
6	3	93
*	3	*

"Missing values" are entered for both the dependent and the regressor variable.

***** ANALYSIS OF VARIANCE *****

VARIATE: X ← The regression variable, X: estimated missing values are incorporated, just as in Data Sets 2-5.

SOURCE OF VARIATION	DF(MV)	SS	SS%	MS	VR
UNITS STRATUM					
A	2	8.000	68.29	4.000	2.667
RESIDUAL	4(2)	6.000	51.22	1.500	
TOTAL	6	14.000	119.51	2.333	
GRAND TOTAL	6	14.000	119.51		

ESTIMATED GRAND MEAN	3.67
TOTAL NUMBER OF OBSERVATIONS	9
NUMBER OF MISSING VALUES	2
MAXIMUM NUMBER OF ITERATIONS	2

UNIT NUMBER	ESTIMATED VALUE
6	3.00
9	5.00

***** ANALYSIS OF VARIANCE *****

VARIATE: Y ← Independent variable, Y: similar calculations, unrelated to X.

SOURCE OF VARIATION	DF(MV)	SS	SS%	MS	VR
UNITS STRATUM					
A	2	402.00	102.55	201.00	9.805
RESIDUAL	4(2)	82.00	20.92	20.50	
TOTAL	6	484.00	123.47	80.67	
GRAND TOTAL	6	484.00	123.47		

ESTIMATED GRAND MEAN	80.0
TOTAL NUMBER OF OBSERVATIONS	9
NUMBER OF MISSING VALUES	2
MAXIMUM NUMBER OF ITERATIONS	2

UNIT NUMBER	ESTIMATED VALUE
6	78.0
9	89.0

***** ANALYSIS OF VARIANCE *****
(ADJUSTED FOR COVARIATES)

VARIATE: Y

SOURCE OF VARIATION		DF(MV)	SS	SS%	MS	VR	COV EF
UNITS STRATUM							
A		2	171.43	43.73	85.71	3.194	0.600
COVARIATES	$R(b\mid\mu,\alpha)$	1	1.50	0.38	1.50	0.056	
RESIDUAL	SSE	3(2)	80.50	20.54	26.83		0.764
TOTAL		6	253.43	64.65	42.24		
GRAND TOTAL		6	253.43	64.65			

In the treatment line, COV EF is <u>cov</u>ariance <u>e</u>fficiency <u>f</u>actor

$$= 1/\left[1 + \frac{\text{MS(Treatment for covariate)}}{\text{SSE(for covariate)}}\right]$$

$$= \frac{1}{1 + \frac{2}{3}} = \frac{3}{5} = 0.6 .$$

In the error line, COV EF is

$$= \frac{\text{MS Residual without covariate}}{\text{MS Residual with covariate}} = \frac{20.50}{26.83}$$

ESTIMATED GRAND MEAN	80.0
TOTAL NUMBER OF OBSERVATIONS	9
NUMBER OF MISSING VALUES	2
MAXIMUM NUMBER OF ITERATIONS	2

UNIT NUMBER	ESTIMATED VALUE
6	78.3
9	88.3

The covariance analysis is done of Y on X, including the EMV's of both X and Y. Because the model is $y_{ij} = \mu + \alpha_i + bx_{ij} + e_{ij}$, b is estimated as the pooled within-group regression coefficient, see (31) p. 349, (33) p. 350, and (42) p. 353 in LM. And because of this pooled within-group estimation, the estimated missing values (group means) do not affect the estimation and so it is correct. So also is the sum of squares due to the covariate, and for error.

***** COVARIANCE REGRESSIONS *****

COVARIATE	COEFFICIENT	SE
UNITS STRATUM		
X	0.5	2.11

$\hat{b}$ of (42), LM p. 353.

***** TABLES OF MEANS *****
(ADJUSTED FOR COVARIATES)

VARIATE: Y

GRAND MEAN 80.0

A	1	2	3
	73.3	78.3	88.3

The <u>estimated</u> grand mean is not adjusted for covariates. These are means adjusted for covariates expressed as deviations from their means and are

$$\left\{\bar{y}_{i.} - \hat{b}(\bar{x}_{i.} - \bar{x}_{..})\right\}_{i=1\cdots 3} = \begin{bmatrix} 73 - \frac{1}{2}(3 - \frac{11}{3}) \\ 78 - \frac{1}{2}(3 - \frac{11}{3}) \\ 89 - \frac{1}{2}(5 - \frac{11}{3}) \end{bmatrix} = \begin{bmatrix} 73\frac{1}{3} \\ 78\frac{1}{3} \\ 88\frac{1}{3} \end{bmatrix}$$

***** STANDARD ERRORS OF DIFFERENCES OF MEANS *****

TABLE	A
REP	3
SED	5.46

(NOT ADJUSTED FOR MISSING VALUES)

***** STRATUM STANDARD ERRORS AND COEFFICIENTS OF VARIATION *****

STRATUM	DF	SE	CV%
UNITS	3	5.18	6.5

```
GENSTAT V     MARK 3.09
(C)1976 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STATION)
```



```
'REFERENCE' SET_7
'CAPTION'
            ''
               DATA SET 7
               UNBALANCED DATA. TWO WAY CROSSED CLASSIFICATION
               WITH 1 COVARIATE
                  LINEAR MODELS BY S R SEARLE PAGE 375
            ''
'UNITS'        $ 48
'READ/S' Y,X
'FACTOR'       A $ 3
  :            B $ 4
'GENERATE'  A,B,4
'PRINT/P'   A,B,X,Y  $10.0
'TREATMENTS' A*B
'COVARIATES'  X
'ANOVA'        Y
'RUN'
```

```
DATA SET 7
UNBALANCED DATA. TWO WAY CROSSED CLASSIFICATION
WITH 1 COVARIATE
   LINEAR MODELS BY S R SEARLE PAGE 375

Y  0  MNMINMAX    11.0000    6.0000    16.0000   48  VALUES   30 MISSING
X  0  MNMINMAX     4.7778    2.0000     8.0000   48  VALUES   30 MISSING
```

A	B	X	Y
1	1	2	8
1	1	4	13
1	1	3	9
1	1	*	*
1	2	*	*
1	2	*	*
1	2	*	*
1	2	*	*
1	3	7	12
1	3	*	*
1	3	*	*
1	3	*	*
1	4	3	7
1	4	5	11
1	4	*	*
1	4	*	*
2	1	5	6
2	1	3	12
2	1	*	*
2	1	*	*
2	2	6	12
2	2	4	14
2	2	*	*
2	2	*	*
2	3	*	*
2	3	*	*
2	3	*	*
2	3	*	*
2	4	*	*
2	4	*	*
2	4	*	*
2	4	*	*
3	1	*	*
3	1	*	*
3	1	*	*
3	1	*	*
3	2	6	9
3	2	2	7
3	2	*	*
3	2	*	*
3	3	6	14
3	3	8	16
3	3	*	*
3	3	*	*
3	4	4	10
3	4	6	14
3	4	5	11
3	4	7	13

27

***** ANALYSIS OF VARIANCE *****

VARIATE: X

SOURCE OF VARIATION	DF(MV)	SS	SS%	MS	VR
UNITS STRATUM					
A	2	13.468	25.36	6.734	2.928
B	3	99.607	187.54	33.202	14.435
A.B	2(4)	2.632	4.96	1.316	0.572
RESIDUAL	10(26)	23.001	43.31	2.300	
TOTAL	17	138.707	261.16	8.159	
GRAND TOTAL	17	138.707	261.16		

ESTIMATED GRAND MEAN	5.02
TOTAL NUMBER OF OBSERVATIONS	48
NUMBER OF MISSING VALUES	30
MAXIMUM NUMBER OF ITERATIONS	10

UNIT NUMBER	ESTIMATED VALUE
4	3.00
5	3.54
6	3.54
7	3.54
8	3.54
10	6.96
11	6.96
12	6.96
15	4.00
16	4.00
19	4.00
20	4.00
23	5.00
24	5.00
25	7.97
26	7.97
27	7.97
28	7.97
29	5.74
30	5.74
31	5.74
32	5.74
33	3.47
34	3.47
35	3.47
36	3.47
39	4.00
40	4.00
43	7.00
44	7.00

***** ANALYSIS OF VARIANCE *****

VARIATE: Y

SOURCE OF VARIATION	DF(MV)	SS	SS%	MS	VR
UNITS STRATUM					
A	2	52.488	38.03	26.244	4.686
B	3	158.320	114.72	52.773	9.423
A.B	2(4)	65.101	47.17	32.551	5.812
RESIDUAL	10(26)	56.002	40.58	5.600	
TOTAL	17	331.911	240.52	19.524	
GRAND TOTAL	17	331.911	240.52		

ESTIMATED GRAND MEAN	11.15
TOTAL NUMBER OF OBSERVATIONS	48
NUMBER OF MISSING VALUES	30
MAXIMUM NUMBER OF ITERATIONS	7

UNIT NUMBER	ESTIMATED VALUE
4	10.00
5	8.62
6	8.62
7	8.62
8	8.62
10	11.94
11	11.94
12	11.94
15	9.00
16	9.00
19	9.00
20	9.00
23	13.00
24	13.00
25	15.39
26	15.39
27	15.39
28	15.39
29	12.42
30	12.42
31	12.42
32	12.42
33	9.46
34	9.46
35	9.46
36	9.46
39	8.00
40	8.00
43	15.00
44	15.00

***** ANALYSIS OF VARIANCE *****
(ADJUSTED FOR COVARIATES)

VARIATE: Y

SOURCE OF VARIATION		DF(MV)	SS	SS%	MS	VR	COV EF
UNITS STRATUM							
A		2	16.861	12.22	8.431	1.560	0.774
B		3	10.441	7.57	3.480	0.644	0.409
A.B		2(4)	59.161	42.87	29.580	5.472	0.946
COVARIATES	$R(b\|\mu,\alpha,\beta,\gamma)$	1	7.349	5.33	7.349	1.360	
RESIDUAL	SSE	9(26)	48.653	35.26	5.406		1.036
TOTAL		17	142.464	103.23	8.380		
GRAND TOTAL		17	142.464	103.23			

ESTIMATED GRAND MEAN 11.15
TOTAL NUMBER OF OBSERVATIONS 48
NUMBER OF MISSING VALUES 30
MAXIMUM NUMBER OF ITERATIONS 7

UNIT NUMBER	ESTIMATED VALUE
4	11.14
5	9.45
6	9.45
7	9.45
8	9.45
10	10.84
11	10.84
12	10.84
15	9.57
16	9.57
19	9.57
20	9.57
23	13.01
24	13.01
25	13.72
26	13.72
27	13.72
28	13.72
29	12.02
30	12.02
31	12.02
32	12.02
33	10.33
34	10.33
35	10.33
36	10.33
39	8.58
40	8.58
43	13.88
44	13.88

Covariance model here is

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + bx_{ijk} + e_{ijk} .$$

It is a pooled within cell regression.

```
***** COVARIANCE REGRESSIONS *****

  COVARIATE                        COEFFICIENT          SE

*UNITS* STRATUM
  X                                       0.57       0.485


***** TABLES OF MEANS *****
(ADJUSTED FOR COVARIATES)

VARIATE: Y


    GRAND MEAN    11.15

             A         1         2         3
               10.25     12.08     11.13


             B         1         2         3         4
               10.35     10.34     12.82     11.11


             B         1         2         3         4
             A
             1     11.14      9.45     10.85      9.57
             2      9.57     13.01     13.72     12.02
             3     10.33      8.57     13.88     11.73


***** STANDARD ERRORS OF DIFFERENCES OF MEANS *****

TABLE                         A             B             A
                                                          B
-------------------------------------------------------------
REP                          16            12             4
SED                       0.935         1.484         1.999

(NOT ADJUSTED FOR MISSING VALUES)


***** STRATUM STANDARD ERRORS AND COEFFICIENTS OF VARIATION *****

STRATUM                      DF             SE          CV%

*UNITS*                       9          2.325         20.8



  27  'REFERENCE' SET_2
 ******* END OF         SET_7    AT LINE   17    USED 626    LEFT 7374
```

GENSTAT ANOVA: Bartlett's covariance technique for missing data.

Data Set 2

This method uses covariance on dummy variables for missing values. It begins with a model for both existing and "missing" data, using unknown values $-b_{ij}$ for the ij'th missing datum.

$$\underset{\sim}{y} = \begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ -b_{21} \\ -b_{22} \\ y_{23} \\ y_{31} \\ -b_{32} \\ y_{33} \\ y_{41} \\ y_{42} \\ y_{43} \end{bmatrix} = \left[\begin{array}{c|cccc|ccc} 1 & 1 & . & . & . & 1 & . & . \\ 1 & 1 & . & . & . & . & 1 & . \\ 1 & 1 & . & . & . & . & . & 1 \\ 1 & . & 1 & . & . & 1 & . & . \\ 1 & . & 1 & . & . & . & 1 & . \\ 1 & . & 1 & . & . & . & . & 1 \\ 1 & . & . & 1 & . & 1 & . & . \\ 1 & . & . & 1 & . & . & 1 & . \\ 1 & . & . & 1 & . & . & . & 1 \\ 1 & . & . & . & 1 & 1 & . & . \\ 1 & . & . & . & 1 & . & 1 & . \\ 1 & . & . & . & 1 & . & . & 1 \end{array}\right] \left[\begin{array}{c} \mu \\ \hline \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \hline \beta_1 \\ \beta_2 \\ \beta_3 \end{array}\right] + \underset{\sim}{e} = \underset{\approx}{X}\underset{\sim}{\alpha} + \underset{\sim}{e} .$$

```
GENSTAT V     MARK 3.09
(C)1976 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STATION)

    'REFERENCE' SET_2
    'CAPTION'
            ''
                   DATA SET 2
                   UNBALANCED DATA, NO INTERACTION, N(I,J) = 0 OR 1
                   LINEAR MODELS BY S R SEARLE  PAGE 262
            ''
    'UNITS'        $ 12
    'READ/P'       Y
    'NAME'         LEVELSA = X,Y,Z,W
      :            LEVELSB = A,B,C
    'FACTOR'       A $ LEVELSA
      :            B $ LEVELSB
    'GENERATE'     A,B
    'PRINT/P'      A,B,Y $10.0
    ''SET UP A COVARIATE CORRESPONDING TO EACH MISSING VALUE
       WITH A 1 CORRESPONDING TO THE MISSING VALUE ZERO ELSEWHERE''
    'VARIATE'      M1 = 3(0),1,8(0)
      :            M2 = 4(0),1,7(0)
      :            M3 = 7(0),1,4(0)
    'TREATMENTS'    A + B
    'COVARIATES' M1,M2,M3
    'ANOVA'        Y
    'RUN'
```

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

   Y  0  MNMINMAX       9.0000      0.0000      24.0000    12  VALUES    0 MISSING
```

A	B	Y
X	A	18
X	B	12
X	C	24
Y	A	0
Y	B	0
Y	C	9
Z	A	3
Z	B	0
Z	C	15
W	A	6
W	B	3
W	C	18

This is $\underset{\sim}{y}_0$, i.e., the data vector with zero entered for every "missing value" to make the data balanced.

Analysis of the unbalanced data is now reformulated as a balanced data covariance analysis, based on balanced data $\underset{\sim}{y}_0$, that includes 0's for the missing data. There is one element in $\underset{\sim}{b}$ for each missing datum and $\underset{\sim}{Z}$ is null except for a single unity in each position corresponding to a missing datum.

Note, the GENSTAT manual Part 1, Ch. 6, Sec. 7, p. 2 <u>fails to</u> indicate the need to set the missing values to zero for this analysis.

The data vector is the sum of two vectors, the existing and the missing:

$$\underset{\sim}{y} = \overset{\text{data}}{\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ -b_{21} \\ -b_{22} \\ y_{23} \\ y_{31} \\ -b_{32} \\ y_{33} \\ y_{41} \\ y_{42} \\ y_{43} \end{bmatrix}} = \overset{\text{existing}}{\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ 0 \\ 0 \\ y_{23} \\ y_{31} \\ 0 \\ y_{33} \\ y_{41} \\ y_{42} \\ y_{43} \end{bmatrix}} + \overset{\text{missing}}{\begin{bmatrix} 0 \\ 0 \\ 0 \\ -b_{21} \\ -b_{22} \\ 0 \\ 0 \\ -b_{32} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}} = \underset{\sim}{y}_0 - \begin{bmatrix} . & . & . \\ . & . & . \\ . & . & . \\ 1 & . & . \\ . & 1 & . \\ . & . & . \\ . & . & . \\ . & . & 1 \\ . & . & . \\ . & . & . \\ . & . & . \\ . & . & . \end{bmatrix} \begin{bmatrix} b_{21} \\ b_{22} \\ b_{32} \end{bmatrix} = \underset{\sim}{y}_0 - \underset{\approx}{Z}\underset{\sim}{b} .$$

Hence, $\underset{\sim}{y} = \underset{\sim}{y}_0 - \underset{\approx}{Z}\underset{\sim}{b} = \underset{\approx}{X}\underset{\sim}{\alpha} + \underset{\sim}{e}$ and so, because $\underset{\sim}{b}$ is unknown, we write

$$\underset{\sim}{y}_0 = \underset{\approx}{X}\underset{\sim}{\alpha} + \underset{\approx}{Z}\underset{\sim}{b} + \underset{\sim}{e} .$$

***** ANALYSIS OF VARIANCE *****

VARIATE: M1

SOURCE OF VARIATION	DF	SS	SS%	MS	VR
UNITS STRATUM					
A	3	0.25000	27.27	0.08333	1.000
B	2	0.16667	18.18	0.08333	1.000
RESIDUAL	6	0.50000	54.55	0.08333	
TOTAL	11	0.91667	100.00	0.08333	
GRAND TOTAL	11	0.91667	100.00		

GRAND MEAN 0.083
TOTAL NUMBER OF OBSERVATIONS 12

In this example $\underset{\sim}{Z}$ has 3 columns, of order 12. Each is treated as a covariable — with

11 observations of zero and 1 of 1 .

For example, for M1 the "data" are:

$$\begin{matrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix}$$

All mean squares are $\frac{1}{ab} = \frac{1}{12} = .0833$.

Proof:

$$MSA = \frac{1}{a-1}\left[\frac{(a-1)0^2}{b} + \frac{1^2}{b} - \frac{1^2}{ab}\right] = \frac{a-1}{(a-1)ab} = \frac{1}{ab}$$

$$MSB = \frac{1}{b-1}\left[\frac{(b-1)0^2}{a} + \frac{1^2}{a} - \frac{1^2}{ab}\right] = \frac{1}{ab}$$

$$MSE = \frac{1}{(a-1)(b-1)}\left[1 - \frac{1^2}{ab} - \frac{a-1}{ab} - \frac{b-1}{ab}\right] = \frac{ab-a-b+1}{(a-1)(b-1)ab} = \frac{1}{ab} .$$

***** ANALYSIS OF VARIANCE *****

VARIATE: M2

SOURCE OF VARIATION	DF	SS	SS%	MS	VR
UNITS STRATUM					
A	3	0.25000	27.27	0.08333	1.000
B	2	0.16667	18.18	0.08333	1.000
RESIDUAL	6	0.50000	54.55	0.08333	
TOTAL	11	0.91667	100.00	0.08333	
GRAND TOTAL	11	0.91667	100.00		

GRAND MEAN 0.083
TOTAL NUMBER OF OBSERVATIONS 12

Ditto

***** ANALYSIS OF VARIANCE *****

VARIATE: M3

SOURCE OF VARIATION	DF	SS	SS%	MS	VR
UNITS STRATUM					
A	3	0.25000	27.27	0.08333	1.000
B	2	0.16667	18.18	0.08333	1.000
RESIDUAL	6	0.50000	54.55	0.08333	
TOTAL	11	0.91667	100.00	0.08333	
GRAND TOTAL	11	0.91667	100.00		

GRAND MEAN 0.083
TOTAL NUMBER OF OBSERVATIONS 12

Ditto

```
***** ANALYSIS OF VARIANCE *****

VARIATE: Y

SOURCE OF VARIATION              DF        SS      SS%        MS        VR

*UNITS* STRATUM
  A                               3   378.000    50.00   126.000    33.600
  B                               2   355.500    47.02   177.750    47.400
  RESIDUAL                        6    22.500     2.98     3.750
TOTAL                            11   756.000   100.00    68.727

GRAND TOTAL                      11   756.000   100.00

GRAND MEAN                          9.00
TOTAL NUMBER OF OBSERVATIONS          12
```

Although no account of the covariates is taken here, this is not the same as GENSTAT ANOVA without the covariates. This analysis is balanced, with "observations" of 0 for what were previously missing values. So it is an analysis of variance of the data $\underset{\sim}{y}_0$:

18	12	24	54	(3)	18
0	0	9	9	(3)	3
3	0	15	18	(3)	6
6	3	18	27	(3)	9
27	15	66	108		
(4)	(4)	(4)		(12)	
$6\frac{3}{4}$	$3\frac{3}{4}$	$16\frac{1}{2}$			9

$$
\begin{aligned}
SST &= && = 1728 \\
SSM &= 12(9^2) && = 972 \\
SSA &= 3(18^2 + 3^2 + 6^2 + 9^2) && = 1350 \\
SST_m &= 1728 - 972 && = 756 \\
SSA_m &= 1350 - 972 && = 378 \\
SSB &= (27^2 + 15^2 + 66^2)/4 && = 1327.5 \\
SSB_m &= 1327.5 - 972 && = 355.5
\end{aligned}
$$

***** ANALYSIS OF VARIANCE *****
(ADJUSTED FOR COVARIATES)

VARIATE: Y

SOURCE OF VARIATION		DF	SS	SS%	MS	VR	COV EF
UNITS STRATUM							
A	$R(\alpha\|\mu,\beta)$	3	271.499	35.91	90.500	22.625	0.574
B	$R(\beta\|\mu,\alpha)$	2	258.000	34.13	129.000	32.250	0.603
COVARIATES		3	10.500	1.39	3.500	0.875	
RESIDUAL	SSE	3	12.000	1.59	4.000		0.937
TOTAL		11	551.999	73.02	50.182		
GRAND TOTAL		11	551.999	73.02			

Although $R(\alpha|\mu,\beta)$ and $R(\beta|\mu,\alpha)$ are given correctly, $R(\alpha|\mu)$ and $R(\beta|\mu)$ cannot be derived. SSE is also correct. The sum of squares for covariates has no meaning. Nor does the total.

GRAND MEAN 9.00
TOTAL NUMBER OF OBSERVATIONS 12

***** COVARIANCE REGRESSIONS *****

COVARIATE	COEFFICIENT	SE	CORRELATION MATRIX M1	M2	M3
UNITS STRATUM					
M1	$\hat{b}_1 = \hat{b}_{21} = 1.0$	3.27	1.000		
M2	$\hat{b}_2 = \hat{b}_{22} = 5.0$	3.42	0.478	1.000	
M3	$\hat{b}_3 = \hat{b}_{32} = -0.0$	3.00	0.000	0.293	1.000

These coefficients multiplied by -1 are the EMV's, $(-\hat{b}_1, -\hat{b}_2, -\hat{b}_3) = (-1, -5, 0)$; they are identical to the EMV's derived both by minimizing the SSE (LM, p. 362) and by the GENSTAT iterative technique within rounding error (see GENSTAT ANOVA output). But, in contrast to that, Bartlett's covariance procedure does not put the EMV's into the missing cells; it carries out the analysis of covariance.

***** TABLES OF MEANS *****
(ADJUSTED FOR COVARIATES)

VARIATE: Y

GRAND MEAN 9.00

A	X	Y	Z	W
	18.50	1.50	6.50	9.50

B	A	B	C
	7.00	3.00	17.00

In the covariance analysis, the data vector adjusted for covariates would, in terms of page 32, be $\mathbf{y}_0 - \mathbf{Z}\hat{\mathbf{b}}$. But GENSTAT always uses covariates expressed as deviations from their means, so that the adjusted data vector is

$$\mathbf{y}_a = \mathbf{y}_0 - (\mathbf{Z} - \tfrac{1}{N}\mathbf{J}\mathbf{Z})\hat{\mathbf{b}}$$

where N is the number of observations (observed + missing). By the nature of $\mathbf{Z}$ (see p. 32) in this missing-value context, it has a single unity in each column and zeros elsewhere, so that if m is the number of missing observations, $\mathbf{J}\mathbf{Z} = \mathbf{J}_{N\times N}\mathbf{Z}_{N\times m} = \mathbf{1}_N\mathbf{1}'_m$. Therefore

$$\mathbf{y}_a = \mathbf{y}_0 - \mathbf{Z}\hat{\mathbf{b}} + \tfrac{1}{N}\mathbf{1}_N\mathbf{1}'_m\hat{\mathbf{b}} = \mathbf{y}_0 + \mathbf{Z}(-\hat{\mathbf{b}}) + (\tfrac{1}{N}\Sigma\hat{b}_i)\mathbf{1}_N \,.$$

The adjusted means output are means for levels of each factor of the adjusted data vector, i.e., of $\mathbf{y}_a$. But $\mathbf{y}_0 + \mathbf{Z}(-\hat{\mathbf{b}})$ in $\mathbf{y}_a$ is the vector of data with EMV's included. Therefore the adjusted means output, with $(\frac{1}{N}\Sigma\hat{b}_i)$ subtracted from each, gives the means for levels of each factor of the data with EMV's included.

Example: $\frac{1}{N}(\hat{b}_1 + \hat{b}_2 + \hat{b}_3) = \frac{1}{12}(1 + 5 + 0) = \frac{1}{2}$.

Then column means for the data with EMV's are $7 - \frac{1}{2} = 6\frac{1}{2}$, $3 - \frac{1}{2} = 2\frac{1}{2}$, $17 - \frac{1}{2} = 16\frac{1}{2}$ as in the table (apart from rounding error) on p. 15.

***** STANDARD ERRORS OF DIFFERENCES OF MEANS *****

TABLE	A	B
REP	3	4
SED	2.156	1.822

***** STRATUM STANDARD ERRORS AND COEFFICIENTS OF VARIATION *****

STRATUM	DF	SE	CV%
UNITS	3	2.000	22.2

```
GENSTAT V    MARK 3.09
(C)1976 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STATION)

'CAPTION'
           ''
             DATA SET 3
             UNBALANCED DATA, TWO WAY CROSSED CLASSIFICATION
                WITH INTERACTION  N(I,J) > 0
                    EXAMPLE FROM BU-608-M BY S R SEARLE
           ''
'UNITS'      $ 12
'READ/P'     Y
'FACTOR'     A $ 2
   :         B $ 3
'GENERATE'   A,B,2
'CAPTION'
           ''
             SET UP COVARIATES FOR EACH MISSING VALUE
           ''
'VARIATE'    M1 = 3(0),1,8(0)
 :           M2 = 5(0),1,6(0)
 :           M3 = 7(0),1,4(0)
  :          M4 = 11(0),1
'PRINT/P'    A,B,Y $10.0
'TREATMENTS' A*B
'COVARIATES' M1,M2,M3,M4
'ANOVA'      Y
'RUN'

    Y  0  MNMINMAX      4.6667      0.0000      12.0000   12  VALUES    0 MISSING

       SET UP COVARIATES FOR EACH MISSING VALUE
```

A	B	Y
1	1	7
1	1	9
1	2	6
1	2	0
1	3	2
1	3	0
2	1	8
2	1	0
2	2	4
2	2	8
2	3	12
2	3	0

***** ANALYSIS OF VARIANCE *****

VARIATE: M1

SOURCE OF VARIATION	DF	SS	SS%	MS	VR
UNITS STRATUM					
A	1	0.08333	9.09	0.08333	1.000
B	2	0.16667	18.18	0.08333	1.000
A.B	2	0.16667	18.18	0.08333	1.000
RESIDUAL	6	0.50000	54.55	0.08333	
TOTAL	11	0.91667	100.00	0.08333	
GRAND TOTAL	11	0.91667	100.00		

GRAND MEAN 0.083
TOTAL NUMBER OF OBSERVATIONS 12

Analysis of variance of covariates.

The first one is 0,0 1,0 0,0
0,0 0,0 0,0

The row sum of squares is

$$\frac{1}{a-1}\left[bn\left(\frac{1}{bn}\right)^2 + (a-1)0^2 - abn\left(\frac{1}{abn}\right)^2\right] = \frac{1}{abn} = \frac{1}{12} = .0833 .$$

***** ANALYSIS OF VARIANCE *****

VARIATE: M2

SOURCE OF VARIATION	DF	SS	SS%	MS	VR
UNITS STRATUM					
A	1	0.08333	9.09	0.08333	1.000
B	2	0.16667	18.18	0.08333	1.000
A.B	2	0.16667	18.18	0.08333	1.000
RESIDUAL	6	0.50000	54.55	0.08333	
TOTAL	11	0.91667	100.00	0.08333	
GRAND TOTAL	11	0.91667	100.00		

GRAND MEAN 0.083
TOTAL NUMBER OF OBSERVATIONS 12

***** ANALYSIS OF VARIANCE *****

VARIATE: M4

SOURCE OF VARIATION	DF	SS	SS%	MS	VR
UNITS STRATUM					
A	1	0.08333	9.09	0.08333	1.000
B	2	0.16667	18.18	0.08333	1.000
A.B	2	0.16667	18.18	0.08333	1.000
RESIDUAL	6	0.50000	54.55	0.08333	
TOTAL	11	0.91667	100.00	0.08333	
GRAND TOTAL	11	0.91667	100.00		

GRAND MEAN 0.083
TOTAL NUMBER OF OBSERVATIONS 12

***** ANALYSIS OF VARIANCE *****

VARIATE: M3

SOURCE OF VARIATION	DF	SS	SS%	MS	VR
UNITS STRATUM					
A	1	0.08333	9.09	0.08333	1.000
B	2	0.16667	18.18	0.08333	1.000
A.B	2	0.16667	18.18	0.08333	1.000
RESIDUAL	6	0.50000	54.55	0.08333	
TOTAL	11	0.91667	100.00	0.08333	
GRAND TOTAL	11	0.91667	100.00		

GRAND MEAN 0.083
TOTAL NUMBER OF OBSERVATIONS 12

***** ANALYSIS OF VARIANCE *****

VARIATE: Y

SOURCE OF VARIATION	DF	SS	SS%	MS	VR
UNITS STRATUM					
A	1	5.33	2.71	5.33	0.239
B	2	12.67	6.44	6.33	0.284
A.B	2	44.67	22.71	22.33	1.000
RESIDUAL	6	134.00	68.14	22.33	
TOTAL	11	196.67	100.00	17.88	
GRAND TOTAL	11	196.67	100.00		

GRAND MEAN 4.7
TOTAL NUMBER OF OBSERVATIONS 12

***** ANALYSIS OF VARIANCE *****
(ADJUSTED FOR COVARIATES)

VARIATE: Y

SOURCE OF VARIATION		DF	SS	SS%	MS	VR	COV EF
UNITS STRATUM							
A	SSA_w	1	20.000	10.17	20.000	4.000	0.600
B	SSB_w	2	5.333	2.71	2.667	0.533	0.600
A.B	$R(\gamma\|\mu,\alpha,\beta)$	2	36.364	18.49	18.182	3.636	0.600
COVARIATES		4	124.000	63.05	31.000	6.200	
RESIDUAL	SSE	2	10.000	5.08	5.000		4.467
TOTAL		11	195.697	99.51	17.791		
GRAND TOTAL		11	195.697	99.51			

Because this example has all cells filled, although with unequal numbers of observations, the sums of squares produced here are, for rows and columns, those of the weighted squares of means analysis, SSA_w and SSB_w (LM, p. 370). $R(\gamma|\mu,\alpha,\beta)$ and SSE are also produced.

GRAND MEAN 4.7
TOTAL NUMBER OF OBSERVATIONS 12

***** COVARIANCE REGRESSIONS *****

COVARIATE	COEFFICIENT	SE	CORRELATION MATRIX M1	M2	M3	M4
UNITS STRATUM						
M1	-6.0	3.16	1.000			
M2	-2.0	3.16	0.000	1.000		
M3	-8.0	3.16	0.000	0.000	1.000	
M4	-12.0	3.16	0.000	0.000	0.000	1.000

EMV's are the cell means 6, 2, 8 and 12. (See p. 17.)

```
***** TABLES OF MEANS *****
(ADJUSTED FOR COVARIATES)

VARIATE: Y

    GRAND MEAN       4.7

             A        1        2
                    3.0      6.3


             B        1        2        3
                    5.7      3.7      4.7


             B        1        2        3
             A
             1      5.7      3.7     -0.3
             2      5.7      3.7      9.7


***** STANDARD ERRORS OF DIFFERENCES OF MEANS *****

TABLE                      A           B           A
                                                   B
--------------------------------------------------------
REP                        6           4           2
SED                     1.67        2.04        2.89

***** STRATUM STANDARD ERRORS AND COEFFICIENTS OF VARIATION *****

STRATUM                     DF            SE         CV%

*UNITS*                      2          2.24        47.9



   80  'REFERENCE' SET_6
  ******* END OF       CAPTION     AT LINE    76     USED 608     LEFT 7392
```

As on page 36, these are means based on

$$\underset{\sim}{y}_0 + (\underset{\sim}{Z} - \tfrac{1}{N}\underset{\sim}{J}\underset{\sim}{Z})(-\underset{\sim}{\hat{b}})$$

$$= \underset{\sim}{y}_0 - \underset{\sim}{Z}\underset{\sim}{\hat{b}} + \tfrac{1}{N}\underset{\sim}{1}(\underset{\sim}{1}'\underset{\sim}{\hat{b}})$$

$$= \underset{\sim}{y}_0 - \underset{\sim}{Z}\underset{\sim}{\hat{b}} + \tfrac{1}{12}(-6 - 2 - 8 - 12)\underset{\sim}{1}$$

$$= \underset{\sim}{y}_0 - \underset{\sim}{Z}\underset{\sim}{\hat{b}} - 2\tfrac{1}{3}\underset{\sim}{1}.$$

Hence, adding $2\frac{1}{3}$ to each output value yields the estimated cell and marginal means; e.g., 9.7, which is $9\frac{2}{3}$, becomes $9\frac{2}{3} + 2\frac{1}{3} = 12$ as on page 17.

This analysis cannot be done on Data Sets 4, 5 and 7 because they have whole cells missing. (So does Data Set 2, but that has a no-interaction model.)

```
GENSTAT V    MARK 3.09
(C)1976 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STATION)

'REFERENCE' SET_6
'CAPTION'
            ''
                 DATA SET 6
                 UNBALANCED DATA, 1 WAY CLASSIFICATION
                 WITH 1 COVARIATE
                 LINEAR MODELS BY S R SEARLE PAGE 353
            ''
'UNITS'      $ 9
'FACTOR'     A $ 3
'READ/P'     A,Y,X
'PRINT/P'    X,A,Y $ 10.0
''SET UP COVARIATES FOR MISSING DATA''
'VARIATE'    M(1) = 5(0),1,3(0)
  :          M(2) = 8(0),1
'TREATMENTS'    A
'COVARIATES'    X,M(1,2)
'ANOVA'      Y
'RUN'


          DATA SET 6
          UNBALANCED DATA, 1 WAY CLASSIFICATION
          WITH 1 COVARIATE
          LINEAR MODELS BY S R SEARLE PAGE 353

   Y  0  MNMINMAX      61.4444      0.0000      93.0000     9  VALUES    0 MISSING
   X  0  MNMINMAX       2.7778      0.0000       6.0000     9  VALUES    0 MISSING


     X          A          Y
     3          1         74
     4          1         68
     2          1         77
     2          2         76
     4          2         80
     0 <-----   2          0 <---- 0's inserted
     4          3         85       for missing
     6          3         93
     0 <-----   3          0 <---- values.
```

Missing data in $\underset{\sim}{y}$ and $\underset{\sim}{x}$ are given values 0, yielding $\underset{\sim}{y}_0$ and $\underset{\sim}{x}_0$ in the same way as $\underset{\sim}{y}_0$ was derived in Data Set 2. This extends Bartlett's covariance technique to covariates $\underset{\sim}{x}$, as well as the variable $\underset{\sim}{y}$.

***** ANALYSIS OF VARIANCE *****

VARIATE: X

SOURCE OF VARIATION	DF	SS	SS%	MS	VR
UNITS STRATUM					
A	2	2.889	9.15	1.444	0.302
RESIDUAL	6	28.667	90.85	4.778	
TOTAL	8	31.556	100.00	3.944	
GRAND TOTAL	8	31.556	100.00		

GRAND MEAN 2.78
TOTAL NUMBER OF OBSERVATIONS 9

This is the analysis of $\underset{\sim}{x}_0$, with 0's in place of missing observations.

	3	2	4	
	4	4	6	
	2	0	0	
Means	3	2	$3\frac{1}{3}$	$2\frac{7}{9}$

***** ANALYSIS OF VARIANCE *****

VARIATE: M(1)

SOURCE OF VARIATION	DF	SS	SS%	MS	VR
UNITS STRATUM					
A	2	0.2222	25.00	0.1111	1.000
RESIDUAL	6	0.6667	75.00	0.1111	
TOTAL	8	0.8889	100.00	0.1111	
GRAND TOTAL	8	0.8889	100.00		

GRAND MEAN 0.11
TOTAL NUMBER OF OBSERVATIONS 9

	0	0	0	
	0	0	0	
	0	1	0	
Means	0	$\frac{1}{3}$	0	$\frac{1}{9}$

$R(\mu) = 9\left(\frac{1}{9}\right)^2 = \frac{1}{9}$

$R(\alpha|\mu) = 3\left(\frac{1}{3}\right)^2 - 9\left(\frac{1}{9}\right)^2 = \frac{1}{3} - \frac{1}{9} = \frac{2}{9}$

$SST = 1$

$CFM = \frac{1}{9}$

***** ANALYSIS OF VARIANCE *****

VARIATE: M(2)

SOURCE OF VARIATION	DF	SS	SS%	MS	VR
UNITS STRATUM					
A	2	0.2222	25.00	0.1111	1.000
RESIDUAL	6	0.6667	75.00	0.1111	
TOTAL	8	0.8889	100.00	0.1111	
GRAND TOTAL	8	0.8889	100.00		

GRAND MEAN 0.11
TOTAL NUMBER OF OBSERVATIONS 9

```
***** ANALYSIS OF VARIANCE *****

VARIATE: Y

SOURCE OF VARIATION             DF        SS       SS%        MS       VR

*UNITS* STRATUM
  A                              2       682      6.75       341    0.217
  RESIDUAL                       6      9419     93.25      1570
TOTAL                            8     10100    100.00      1263

GRAND TOTAL                      8     10100    100.00

GRAND MEAN                          61
TOTAL NUMBER OF OBSERVATIONS         9
```

The SS and mean are rounded. The default number of decimal places printed for the means is such that the residual (stratum) variance may be printed to 2 significant figures. User may change this using a DESCRIBE statement. The mean is given as 61.4444 in the data summary on the first page.

This is the analysis of $\underset{\sim}{y}_0$, with 0's in place of missing values:

	74	76	85	
	68	80	93	
	77	0	0	
Means $\bar{y}_{0i.}$	73	52	$59\frac{1}{3}$	$61\frac{4}{9}$ $\bar{y}_{0..}$

$R(\mu) = 33,978\frac{7}{9}$

$R(\alpha) = 3\left(73^2 + 52^2 + (59\frac{1}{3})^2\right) = 34,660\frac{3}{9}$

$R(\alpha|\mu) = 681\frac{5}{9}$

$SST = 44,079$

$SST_m = 44,079 - 33,978\frac{7}{9} = 10,100\frac{2}{9}$

$SSE = 9,418\frac{6}{9}$

***** ANALYSIS OF VARIANCE *****
(ADJUSTED FOR COVARIATES)

VARIATE: Y

SOURCE OF VARIATION		DF	SS	SS%	MS	VR	COV EF
UNITS STRATUM							
A	$R(\alpha\|\mu,b)$	2	153.66	1.52	76.83	2.863	0.476
COVARIATES		3	9338.17	92.46	3112.72	116.002	
RESIDUAL	SSE	3	80.50	0.80	26.83		58.501
TOTAL		8	9572.33	94.77	1196.54		
GRAND TOTAL		8	9572.33	94.77			

GRAND MEAN 61
TOTAL NUMBER OF OBSERVATIONS 9

This procedure yields the correct values $\bar{y}$, $R(\alpha|\mu,b)$, SSE and $\hat{b}$ for unbalanced data.

ditto: mean $61\frac{4}{9}$ rounded to 61.

***** COVARIANCE REGRESSIONS *****

COVARIATE		COEFFICIENT	SE	CORRELATION MATRIX X	M(1)	M(2)
UNITS STRATUM						
X	$\hat{b}$	0.5	2.11	1.000		
M(1)	$\hat{b}_1$	-76.5	8.97	0.707	1.000	
M(2)	$\hat{b}_2$	-86	12.3	0.857	0.606	1.000

$\hat{b} = 0.5$ is correct.

***** TABLES OF MEANS *****
(ADJUSTED FOR COVARIATES)

VARIATE: Y

GRAND MEAN 61

A	1	2	3
	55	60	70

These are means adjusted for covariate, $\underset{\sim}{x}_0$, and for missing values, M_1 and M_2 .

The grand mean is unadjusted; its value is $61\frac{4}{9}$ rounded to 61.

The A adjusted means are not useful as they involve adjustment by $\underset{\sim}{x}_0$, the covariate with 0's for missing values; the calculation is

$$\left\{\bar{y}_{0i.} - \hat{b}(\bar{x}_{0i.} - \bar{x}_{0..}) - \sum_{k=1}^{2}\hat{b}_k(\bar{m}_{ki.} - \bar{m}_{k..})\right\}_{i=1\cdots 3}$$

$$= \begin{bmatrix} 73 \\ 52 \\ 59\frac{1}{3} \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 3 - 2\frac{7}{9} \\ 2 - 2\frac{7}{9} \\ 3\frac{1}{3} - 2\frac{7}{9} \end{bmatrix} + 76.5\begin{bmatrix} 0 - \frac{1}{9} \\ \frac{1}{3} - \frac{1}{9} \\ 0 - \frac{1}{9} \end{bmatrix} + 86\begin{bmatrix} 0 - \frac{1}{9} \\ 0 - \frac{1}{9} \\ \frac{1}{3} - \frac{1}{9} \end{bmatrix} = \begin{bmatrix} 54\frac{5}{6} \\ 59\frac{5}{6} \\ 69\frac{2}{3} \end{bmatrix} .$$

In output these are rounded to whole numbers. The mean of these adjusted means is $\bar{y}_{..}$, as expected, $61\frac{4}{9}$.

***** STANDARD ERRORS OF DIFFERENCES OF MEANS *****

TABLE	A
REP	3
SED	6.1

***** STRATUM STANDARD ERRORS AND COEFFICIENTS OF VARIATION *****

STRATUM	DF	SE	CV%
UNITS	3	5.2	8.4

102 'CLOSE'

```
GENSTAT V    MARK 3.09
(C)1976 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STATION)

'REFERENCE' SET_2
'CAPTION'
              ''
                  DATA SET 2
                  UNBALANCED DATA, NO INTERACTION, N(I,J) = 0 OR 1
                  LINEAR MODELS BY S R SEARLE  PAGE 262
              ''
'UNITS'       $ 12
'READ/P'      Y
'NAME'        LEVELSA = X,Y,Z,W
  :           LEVELSB = A,B,C
'FACTOR'      A $ LEVELSA
  :           B $ LEVELSB
'GENERATE'    A,B
'PRINT/P'     A,B,Y $10.0
'PAGE'
'CAPTION'
              ''
                  UNBALANCED ANALYSIS USING REGRESSION
              ''
'TERMS/PRINT=SC'   A,B,Y
'Y'                Y
'PAGE'
'ADD/PRINT=Z, ANDEV=I'       A
'ADD/PRINT=ACIVUX, ANDEV=T'  B     1'st order of fitting
'DROP/PRINT=Z'               A,B
'ADD/PRINT=Z,ANDEV=I'        B
'ADD/PRINT=Z,ANDEV=T'        A     2'nd order of fitting
'RUN'
```

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

Y  0  MNMINMAX     12.0000     3.0000     24.0000    12  VALUES    3 MISSING
```

A	B	Y
X	A	18
X	B	12
X	C	24
Y	A	*
Y	B	*
Y	C	9
Z	A	3
Z	B	*
Z	C	15
W	A	6
W	B	3
W	C	18

Units with missing values (designated *) are omitted from regression calculations.

```
21..............................................................................................

***** SUMS OF SQUARES AND PRODUCTS MATRIX *****

        DF =   8

Y            1   4.3200E  2
A 1          2   1.8000E  1   2.0000E  0
A 2          3  -3.0000E  0  -3.3333E -1   8.8889E -1
A 3          4  -6.0000E  0  -6.6667E -1  -2.2222E -1   1.5556E  0
B 1          5  -9.0000E  0   1.3878E-17  -3.3333E -1   3.3333E -1   2.0000E  0
B 2          6  -9.0000E  0   3.3333E -1  -2.2222E -1  -4.4444E -1  -6.6667E -1   1.5556E  0
MEAN         7   1.2000E  1   3.3333E -1   1.1111E -1   2.2222E -1   3.3333E -1   2.2222E -1   9.0000E  0

                      1            2            3            4            5            6            7
```

This matrix (excluding its last row) is the corrected SSP matrix of data which are columns of $\underset{\sim}{M}$, the matrix of $\underset{\sim}{y}$, and of $\underset{\sim}{X}$ with columns corresponding to u, α_4 and β_3 omitted.

Example:

$$\underset{\sim}{M} = \begin{array}{c} \begin{array}{cccccc} & & & & \alpha_4 & \beta_3 \end{array} \\ \begin{bmatrix} 18 & 1 & . & . & 1 & . \\ 12 & 1 & . & . & . & 1 \\ 2 & 1 & . & . & . & . \\ 9 & . & 1 & . & . & . \\ 3 & . & . & 1 & 1 & . \\ 15 & . & . & 1 & . & . \\ 6 & . & . & . & 1 & . \\ 3 & . & . & . & . & 1 \\ 18 & . & . & . & . & . \end{bmatrix} \end{array}$$

Means for columns of $\underset{\sim}{M}$.

N

These matrices are requested as optional output by the 'TERMS' statement which creates the corrected SSP matrix with the column means and N appended. Subsequent regression calculations use subsets of this.

```
***** CORRELATION MATRIX *****

        DF =   7

Y            1   1.0000
A 1          2   0.6124   1.0000
A 2          3  -0.1531  -0.2500   1.0000
A 3          4  -0.2315  -0.3780  -0.1890   1.0000
B 1          5  -0.3062   0.0000  -0.2500   0.1890   1.0000
B 2          6  -0.3472   0.1890  -0.1890  -0.2857  -0.3780   1.0000
MEAN         7   0.1925   0.0786   0.0393   0.0594   0.0786   0.0594   1.0000

                      1        2        3        4        5        6        7
```

Why does this differ from DF = 8 above?

No indication as to what has happened to A4 and B3.

These "correlations" are meaningless; e.g.,

$$.0594 = \frac{.2222}{\sqrt{9(1.555)}} .$$

This line has been deleted in version 4.01.

25...

$$\left[\begin{array}{cccccc} 1 & 1 & . & . & 1 & . \\ 1 & 1 & . & . & . & 1 \\ 1 & 1 & . & . & . & . \\ 1 & . & 1 & . & . & . \\ 1 & . & . & 1 & 1 & . \\ 1 & . & . & 1 & . & . \\ 1 & . & . & . & 1 & . \\ 1 & . & . & . & . & 1 \\ 1 & . & . & . & . & . \end{array}\right]$$

***** REGRESSION ANALYSIS *****

*** INVERSE MATRIX ***

Note that this is CONSTANT, not Y. It is for μ, not y.

This is $(\underset{\sim}{X}'\underset{\sim}{X})^{-1}$ for $\underset{\sim}{X}'\underset{\sim}{X}$ that is the corrected SSP matrix of $[\underset{\sim}{1} \quad \underset{\sim}{X}_1]$ = the matrix above.

		1	2	3	4	5	6
CONSTANT	1	5.8333E -1					
A 1	2	-3.3333E -1	6.6667E -1				
A 2	3	-5.8333E -1	3.3333E -1	1.5833E 0			
A 3	4	-4.1667E -1	3.3333E -1	4.1667E -1	9.1667E -1		
B 1	5	-3.3333E -1	-1.8504E-17	3.3333E -1	6.9389E-17	6.6667E -1	
B 2	6	-4.1667E -1	3.2382E-17	4.1667E -1	2.5000E -1	3.3333E -1	9.1667E -1

*** VARIANCE-COVARIANCE MATRIX ***

This is $\hat{\sigma}^2(\underset{\sim}{X}'\underset{\sim}{X})^{-1}$, the <u>estimated</u> variance-covariance matrix for parameters omitting α_4 and β_3, for $\hat{\sigma}^2$ being the error mean square, MSE; in this case $\hat{\sigma}^2 = 12/3 = 4$.

Y-VARIATE: Y

		1	2	3	4	5	6
CONSTANT	1	2.3333E 0					
A 1	2	-1.3333E 0	2.6667E 0				
A 2	3	-2.3333E 0	1.3333E 0	6.3333E 0			
A 3	4	-1.6667E 0	1.3333E 0	1.6667E 0	3.6667E 0		
B 1	5	-1.3333E 0	-7.4015E-17	1.3333E 0	2.7756E-16	2.6667E 0	
B 2	6	-1.6667E 0	1.2953E-16	1.6667E 0	10.0000E -1	1.3333E 0	3.6667E 0

*** REGRESSION COEFFICIENTS ***

This should be SOLUTION. In non full rank models it is <u>not</u> an estimate of the parameter vector. The corresponding solution vector is $\underset{\sim}{b}^{o\prime}$ = [17 9 -8 -3 0 -10 -14 0].

Y-VARIATE: Y

	ESTIMATE	S.E.	T
CONSTANT	17.0000	1.5275	11.13
A 1	9.0000	1.6330	5.51
A 2	-8.0000	2.5166	-3.18
A 3	-3.0000	1.9149	-1.57
B 1	-10.0000	1.6330	-6.12
B 2	-14.0000	1.9149	-7.31

This is $\sqrt{\text{diagonal terms of } \hat{\sigma}^2(\underset{\sim}{X}'\underset{\sim}{X})^{-1}}$. It is of no use because as the standard error of the terms labeled "Estimate" it is the standard error of a solution value that is not invariant; e.g., for the first element, μ is not estimable and so μ^o is not invariant.

This t-test value is of no use, for the same reason. In contrast, for example, $\alpha_1 - \alpha_2$ is estimable, its b.l.u.e. is $\widehat{\alpha_1 - \alpha_2} = \alpha_1^o - \alpha_2^o = 9 - (-8) = 17$, and the standard error of this is, from $\hat{\sigma}^2(\underset{\sim}{X}'\underset{\sim}{X})^{-1}$,

$$\sqrt{2.666 + 2(1.333) + 6.333} = \sqrt{11\tfrac{2}{3}} = 3.41 .$$

Note that this differs from 1.63 + 2.51 = 4.14.

*** ANALYSIS OF VARIANCE ***

	DF	SS		MS
REGRESSN	5	420.00	SSR_m	84.000
RESIDUAL	3	12.00	SSE	4.000
TOTAL	8	432.00	SST_m	54.000
CHANGE	-2	-258.00		129.000

% VARIANCE ACCOUNTED FOR 92.6

However since $\alpha_4^o = \beta_3^o = 0$, the main effect solutions A_i and B_j are in fact b.l.u.e.'s of $\alpha_i - \alpha_4$ and $\beta_j - \beta_4$, respectively. When the last levels of the factors are controls, the SE and T-tests given are appropriate for testing differences from the controls.

```
*** OBSERVED AND FITTED VALUES ***
              OBSERVED        FITTED      RESIDUAL

        1        18.00         16.00          2.00
        2        12.00         12.00          0.00
        3        24.00         26.00         -2.00
        4         9.00          9.00          0.00
        5         3.00          4.00         -1.00
        6        15.00         14.00          1.00
        7         6.00          7.00         -1.00
        8         3.00          3.00          0.00
        9        18.00         17.00          1.00
```

These are values of the b.l.u.e. $\widehat{\mu + \alpha_i + \beta_j} = \mu^o + \alpha_i^o + \beta_j^o$.

Note: The sum of squares of these values is the SSE: $2^2 + 2^2 + 4(1^2) = 12$.

```
*** SET OF FITTED TERMS ***

CONSTANT
A
B
```

```
*** ANALYSIS OF DEVIANCE ***

Y-VARIATE: Y

 TERMS                          RESIDUAL        CHANGE             MEAN    MN DEV.
                               DF   DEVIANCE   DF   DEVIANCE       CHANGE   RATIO
INITIAL MODEL
 CONSTANT                       8 SST_m 432.000   *          *

MODIFICATIONS TO MODEL
+A                              5     270.000   3    162.000      54.000    13.50
+B                              3 SSE  12.000   2    258.000     129.000    32.25
** DENOMINATOR OF RATIO IS (RES. DEV./RES. DF) FROM LINE ABOVE, =       4.000 = MSE
```

For customary regression models the "deviance" is the SSE and the analysis of deviance is formed from the differences of the deviances from increasingly complex models, and so yields the usual analysis of variance.

Ratio of MS's using the smallest MS in the deviance column as denominator. May or may not be appropriate for F-tests.

Mean squares, e.g., MSA = 162/3 = 54.

$R(\alpha|\mu)$

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

```
*** ANALYSIS OF DEVIANCE ***

Y-VARIATE: Y

 TERMS                          RESIDUAL        CHANGE             MEAN    MN DEV.
                               DF   DEVIANCE   DF   DEVIANCE       CHANGE   RATIO
INITIAL MODEL
 CONSTANT                       8     432.000   *          *

MODIFICATIONS TO MODEL
+B                              6     283.500   2    148.500      74.250    18.56
+A                              3      12.000   3    271.500      90.500    22.62
** DENOMINATOR OF RATIO IS (RES. DEV./RES. DF) FROM LINE ABOVE, =       4.000
```

$R(\beta|\mu)$

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

```
  32  'CLOSE'
 ******* END OF         SET_2     AT LINE    21     USED 616     LEFT 7384
```

```
GENSTAT V     MARK 3.09
(C)1976 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STATION)

'REFERENCE' SET_3
'CAPTION'
           ''
              DATA SET 3
              UNBALANCED DATA, TWO WAY CROSSED CLASSIFICATION
                 WITH INTERACTION  N(I,J) > 0
                     EXAMPLE FROM BU-608-M BY S R SEARLE
           ''
'UNITS'      $ 8
'READ/P'     Y
'FACTOR'     A $2 = 4(1,2)
   :         B $ 3 = 1,1,2,3,1,2,2,3
'PRINT/P'    A,B,Y $ 10.0
'TERMS/PRINT=SC'        A*B+Y
'Y'          Y
'CAPTION'
             ''FIT CONSTANT ADD A,B,A.B PARTITIONING SS''
'FIT/PRINT = Z, ANDEV=I'
'ADD/ANDEV=PT,PRINT=XACIVU'  A+B+A.B }
'DROP/PRINT=Z'  A*B                  }
'CAPTION'
             ''FIT CONSTANT ADD B,A,A.B PARTITIONING SS''
'FIT/PRINT = Z, ANDEV=I'
'ADD/ANDEV=PT,PRINT=Z'  B+A+A.B
'RUN'
```

A * B ≡ A + B + A.B

≡ A, and B, and A x B

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

   Y  0  MNMINMAX        7.0000        2.0000        12.0000     8  VALUES     0 MISSING
```

A	B	Y
1	1	7
1	1	9
1	2	6
1	3	2
2	1	8
2	2	4
2	2	8
2	3	12

14..

***** SUMS OF SQUARES AND PRODUCTS MATRIX *****

DF = 7

		1	2	3	4	5	6	7
Y	1	6.6000E 1						
A 1	2	-4.0000E 0	2.0000E 0					
B 1	3	3.0000E 0	5.0000E -1	1.8750E 0				
B 2	4	-3.0000E 0	-5.0000E -1	-1.1250E 0	1.8750E 0			
A 1.B 1	5	2.0000E 0	1.0000E 0	1.2500E 0	-7.5000E -1	1.5000E 0		
A 1.B 2	6	-1.0000E 0	5.0000E -1	-3.7500E -1	6.2500E -1	-2.5000E -1	8.7500E -1	
MEAN	7	7.0000E 0	5.0000E -1	3.7500E -1	3.7500E -1	2.5000E -1	1.2500E -1	8.0000E 0

***** CORRELATION MATRIX *****

DF = 6

		1	2	3	4	5	6	7
Y	1	1.0000						
A 1	2	-0.3482	1.0000					
B 1	3	0.2697	0.2582	1.0000				
B 2	4	-0.2697	-0.2582	-0.6000	1.0000			
A 1.B 1	5	0.2010	0.5774	0.7454	-0.4472	1.0000		
A 1.B 2	6	-0.1316	0.3780	-0.2928	0.4879	-0.2182	1.0000	
MEAN	7	0.3046	0.1250	0.0968	0.0968	0.0722	0.0472	1.0000

FIT CONSTANT ADD A,B,A.B PARTITIONING SS

19...

***** REGRESSION ANALYSIS *****

*** INVERSE MATRIX ***

CONSTANT	1	1.0000E 0					
A 1	2	-1.0000E 0	2.0000E 0				
B 1	3	-1.0000E 0	1.0000E 0	2.0000E 0			
B 2	4	-1.0000E 0	1.0000E 0	1.0000E 0	1.5000E 0		
A 1.B 1	5	1.0000E 0	-2.0000E 0	-2.0000E 0	-1.0000E 0	3.5000E 0	
A 1.B 2	6	1.0000E 0	-2.0000E 0	-1.0000E 0	-1.5000E 0	2.0000E 0	3.5000E 0
		1	2	3	4	5	6

*** VARIANCE-COVARIANCE MATRIX ***

Y-VARIATE: Y

CONSTANT	1	5.0000E 0					
A 1	2	-5.0000E 0	10.0000E 0				
B 1	3	-5.0000E 0	5.0000E 0	10.0000E 0			
B 2	4	-5.0000E 0	5.0000E 0	5.0000E 0	7.5000E 0		
A 1.B 1	5	5.0000E 0	-10.0000E 0	-10.0000E 0	-5.0000E 0	1.7500E 1	
A 1.B 2	6	5.0000E 0	-10.0000E 0	-5.0000E 0	-7.5000E 0	10.0000E 0	1.7500E 1
		1	2	3	4	5	6

*** REGRESSION COEFFICIENTS ***

Y-VARIATE: Y

	ESTIMATE	S.E.	T
CONSTANT	12.0000	2.2361	5.37
A 1	-10.0000	3.1623	-3.16
B 1	-4.0000	3.1623	-1.26
B 2	-6.0000	2.7386	-2.19
A 1.B 1	10.0000	4.1833	2.39
A 1.B 2	10.0000	4.1833	2.39

$$\underset{\sim}{b}^{o'} = [12 \quad \underbrace{-10 \quad 0}_{\alpha's} \quad \underbrace{-4 \quad -6 \quad 0}_{\beta's} \quad \underbrace{10 \quad 10 \quad 0 \quad 0 \quad 0 \quad 0}_{\gamma's}]$$

```
*** ANALYSIS OF VARIANCE ***

                 DF          SS                  MS

  REGRESSN        5      56.000  SSR_m        11.200
  RESIDUAL        2      10.000  SSE           5.000
     TOTAL        7      66.000  SST_m         9.429

    CHANGE       -5     -56.000               11.200

     % VARIANCE ACCOUNTED FOR 47.0

*** OBSERVED AND FITTED VALUES ***
                OBSERVED        FITTED      RESIDUAL

          1         7.00          8.00         -1.00
          2         9.00          8.00          1.00
          3         6.00          6.00         -0.00
          4         2.00          2.00          0.00
          5         8.00          8.00          0.00
          6         4.00          6.00         -2.00
          7         8.00          6.00          2.00
          8        12.00         12.00         -0.00

*** SET OF FITTED TERMS ***

CONSTANT
A
B
A.B


*** ANALYSIS OF DEVIANCE ***

Y-VARIATE: Y

 TERMS                           RESIDUAL       CHANGE            MEAN    MN DEV.
                               DF   DEVIANCE  DF   DEVIANCE      CHANGE    RATIO
INITIAL MODEL
 CONSTANT                       7     66.000   *   SST_m    *

MODIFICATIONS TO MODEL                             SSE
+A                              6     58.000   1      8.000       8.000     1.60
+B                              4     46.364   2     11.636       5.818     1.16
+A.B                            2     10.000   2     36.364      18.182     3.64
** DENOMINATOR OF RATIO IS (RES. DEV./RES. DF) FROM LINE ABOVE, =      5.000
FIT CONSTANT ADD B,A,A.B PARTITIONING SS
```

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

```
*** ANALYSIS OF DEVIANCE ***

Y-VARIATE: Y

 TERMS                           RESIDUAL       CHANGE            MEAN    MN DEV.
                               DF   DEVIANCE  DF   DEVIANCE      CHANGE    RATIO
INITIAL MODEL
 CONSTANT                       7     66.000   *            *

MODIFICATIONS TO MODEL                             SSE
+B                              5     60.000   2      6.000       3.000     0.60
+A                              4     46.364   1     13.636      13.636     2.73
+A.B                            2     10.000   2     36.364      18.182     3.64
** DENOMINATOR OF RATIO IS (RES. DEV./RES. DF) FROM LINE ABOVE, =      5.000
```

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

54

```
GENSTAT V    MARK 3.09
(C)1976 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STATION)

'REFERENCE' SET_4
'CAPTION'
               ''
                    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
               ''
'UNITS'        $ 10
'READ/P'       Y
'FACTOR'       A $2 = 6(1),4(2)
    :          B $ 3 = 1,1,1,2,2,3,1,1,2,2
'PRINT/P'      A,B,Y $ 10.0
'TERMS/PRINT=SC'        A*B+Y
'Y'            Y
'CAPTION'
               ''FIT CONSTANT ADD A,B,A.B PARTITIONING SS''
'FIT/PRINT = Z, ANDEV=I'
'ADD/ANDEV=PT,PRINT=XACIVU'  A+B+A.B
'DROP/PRINT=Z'  A*B
'CAPTION'
               ''FIT CONSTANT ADD B,A,A.B PARTITIONING SS''
'FIT/PRINT = Z, ANDEV=I'
'ADD/ANDEV=PT,PRINT=Z'  B+A+A.B
'RUN'


        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

   Y  0  MNMINMAX        6.5000        2.0000       12.0000    10  VALUES    0 MISSING


    A         B         Y
    1         1         2
    1         1         4
    1         1         6
    1         2         4
    1         2         6
    1         3         5
    2         1        12
    2         1         8
    2         2        11
    2         2         7
```

72...

***** SUMS OF SQUARES AND PRODUCTS MATRIX *****

DF = 9

		1	2	3	4	5	6	7
Y	1	8.8500E 1						
A 1	2	-1.2000E 1	2.4000E 0					
B 1	3	-5.0000E -1	1.3878E-17	2.5000E 0				
B 2	4	2.0000E 0	-4.0000E -1	-2.0000E 0	2.4000E 0			
A 1.B 1	5	-7.5000E 0	1.2000E 0	1.5000E 0	-1.2000E 0	2.1000E 0		
A 1.B 2	6	-3.0000E 0	8.0000E -1	-10.0000E -1	1.2000E 0	-6.0000E -1	1.6000E 0	
MEAN	7	6.5000E 0	6.0000E -1	5.0000E -1	4.0000E -1	3.0000E -1	2.0000E -1	1.0000E 1

***** CORRELATION MATRIX *****

DF = 8

		1	2	3	4	5	6	7
Y	1	1.0000						
A 1	2	-0.8234	1.0000					
B 1	3	-0.0336	0.0000	1.0000				
B 2	4	0.1372	-0.1667	-0.8165	1.0000			
A 1.B 1	5	-0.5501	0.5345	0.6547	-0.5345	1.0000		
A 1.B 2	6	-0.2521	0.4082	-0.5000	0.6124	-0.3273	1.0000	
MEAN	7	0.2185	0.1225	0.1000	0.0816	0.0655	0.0500	1.0000

FIT CONSTANT ADD A,B,A.B PARTITIONING SS

77...

*** LINEAR DEPENDENCE DETECTED WHILE FITTING TERM A.B
1 PARAMETER(S) OF THIS TERM ARE ALIASED.

***** REGRESSION ANALYSIS *****

*** INVERSE MATRIX ***

		1	2	3	4	5	6
CONSTANT	1	2.0000E 0					
A 1	2	-10.0000E -1	10.0000E -1				
B 1	3	-2.0000E 0	10.0000E -1	2.5000E 0			
B 2	4	-1.5000E 0	5.0000E -1	1.5000E 0	1.5000E 0		
A 1.B 1	5	10.0000E -1	-10.0000E -1	-1.5000E 0	-5.0000E -1	1.8333E 0	
A 1.B 2	6	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0

*** VARIANCE-COVARIANCE MATRIX ***

Y-VARIATE: Y

		1	2	3	4	5	6
CONSTANT	1	1.0400E 1					
A 1	2	-5.2000E 0	5.2000E 0				
B 1	3	-1.0400E 1	5.2000E 0	1.3000E 1			
B 2	4	-7.8000E 0	2.6000E 0	7.8000E 0	7.8000E 0		
A 1.B 1	5	5.2000E 0	-5.2000E 0	-7.8000E 0	-2.6000E 0	9.5333E 0	
A 1.B 2	6	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0

*** REGRESSION COEFFICIENTS ***

Y-VARIATE: Y

	ESTIMATE	S.E.	T
CONSTANT	9.0000E 0	3.2249E 0	2.79
A 1	-4.0000E 0	2.2803E 0	-1.75
B 1	10.0000E -1	3.6055E 0	0.28
B 2	-1.2768E-15	2.7928E 0	-0.00
A 1.B 1	-2.0000E 0	3.0876E 0	-0.65
A 1.B 2	0.0000E 0	0.0000E 0	*

$$\underset{\sim}{b}^{o\prime} = [9 \quad \underbrace{-4 \quad 0}_{\alpha's} \quad \underbrace{1 \quad 0 \quad 0}_{\beta's} \quad \underbrace{-2 \quad 0 \quad 0 \quad 0 \quad 0}_{\gamma's}]$$

*** ANALYSIS OF VARIANCE ***

	DF	SS		MS
REGRESSN	4	62.50	SSR_m	15.625
RESIDUAL	5	26.00	SSE	5.200
TOTAL	9	88.50	SST_m	9.833
CHANGE	-4	-62.50		15.625

% VARIANCE ACCOUNTED FOR 47.1

```
*** OBSERVED AND FITTED VALUES ***
          OBSERVED      FITTED     RESIDUAL

  1         2.00         4.00       -2.00
  2         4.00         4.00        0.00
  3         6.00         4.00        2.00
  4         4.00         5.00       -1.00
  5         6.00         5.00        1.00
  6         5.00         5.00       -0.00
  7        12.00        10.00        2.00
  8         8.00        10.00       -2.00
  9        11.00         9.00        2.00
 10         7.00         9.00       -2.00

*** SET OF FITTED TERMS ***

CONSTANT
A
B
A.B


*** ANALYSIS OF DEVIANCE ***

Y-VARIATE: Y

 TERMS                         RESIDUAL        CHANGE              MEAN    MN DEV.
                              DF  DEVIANCE   DF   DEVIANCE         CHANGE   RATIO
INITIAL MODEL
 CONSTANT                      9    88.500    *   SSTm       *

MODIFICATIONS TO MODEL                            SSE
+A                             8    28.500    1     60.000      60.000    16.84
** DENOMINATOR OF RATIO IS (RES. DEV./RES. DF) FROM LINE ABOVE, =        3.562
+B                             6    28.182    2      0.318       0.159     0.04
+A.B                           5    26.000    1      2.182       2.182     0.61
FIT CONSTANT ADD B,A,A.B PARTITIONING SS

 82..................................................................................

*** LINEAR DEPENDENCE DETECTED WHILE FITTING TERM A.B
      1 PARAMETER(S) OF THIS TERM ARE ALIASED.

*** ANALYSIS OF DEVIANCE ***

Y-VARIATE: Y

 TERMS                         RESIDUAL        CHANGE              MEAN    MN DEV.
                              DF  DEVIANCE   DF   DEVIANCE         CHANGE   RATIO
INITIAL MODEL
 CONSTANT                      9    88.500    *              *

MODIFICATIONS TO MODEL
+B                             7    85.200    2      3.300       1.650     0.35
+A                             6    28.182    1     57.018      57.018    12.14
** DENOMINATOR OF RATIO IS (RES. DEV./RES. DF) FROM LINE ABOVE, =        4.697
+A.B                           5    26.000    1      2.182       2.182     0.46


  86  'CLOSE'
 ******* END OF        SET_4     AT LINE    72     USED 624     LEFT 7376
```

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

$16.84 = 60/3.562$

$0.04 = .159/3.562$

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

```
GENSTAT V    MARK 3.09
(C)1976 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STATION)

'REFERENCE' SET_5
'CAPTION'
           ''
                DATA SET 5
                UNBALANCED DATA, TWO WAY CROSSED CLASSIFICATION
                   WITH INTERACTION  SOME N(I,J) = 0
                     LINEAR MODELS BY SRSEARLE PAGE 287
           ''
'UNITS'     $ 18
'READ/P'    Y
'FACTOR'    A $ 3 = 6(1),4(2),8(3)
  :               B  $ 4 = 3(1),3,2(4,1,2,2,3,4,4)
'PRINT/P'   A,B,Y $ 10.0
'TERMS/PRINT=SC'        A*B+Y
'Y'         Y
'CAPTION'
            ''FIT CONSTANT ADD A,B,A.B PARTITIONING SS''
'FIT/PRINT = Z, ANDEV=I'
'ADD/ANDEV=PT,PRINT=XACIVU'  A+B+A.B
'DROP/PRINT=Z'  A*B
'CAPTION'
            ''FIT CONSTANT ADD B,A,A.B PARTITIONING SS''
'FIT/PRINT = Z, ANDEV=I'
'ADD/ANDEV=PT,PRINT=Z'  B+A+A.B
'RUN'


      DATA SET 5
      UNBALANCED DATA, TWO WAY CROSSED CLASSIFICATION
         WITH INTERACTION  SOME N(I,J) = 0
           LINEAR MODELS BY SRSEARLE PAGE 287

   Y  0  MNMINMAX      11.0000      6.0000      16.0000    18  VALUES    0 MISSING
```

A	B	Y
1	1	8
1	1	13
1	1	9
1	3	12
1	4	7
1	4	11
2	1	6
2	1	12
2	2	12
2	2	14
3	2	9
3	2	7
3	3	14
3	3	16
3	4	10
3	4	14
3	4	11
3	4	13

100..

60

***** SUMS OF SQUARES AND PRODUCTS MATRIX *****

DF = 17

Y	1	1.3800E 2						
A 1	2	-6.0000E 0	4.0000E 0					
A 2	3	-5.5511E-16	-1.3333E 0	3.1111E 0				
B 1	4	-7.0000E 0	1.3333E 0	8.8889E -1	3.6111E 0			
B 2	5	-2.0000E 0	-1.3333E 0	1.1111E 0	-1.1111E 0	3.1111E 0		
B 3	6	9.0000E 0	-2.8710E-16	-6.6667E -1	-8.3333E -1	-6.6667E -1	2.5000E 0	
A 1.B 1	7	-3.0000E 0	2.0000E 0	-6.6667E -1	2.1667E 0	-6.6667E -1	-5.0000E -1	2.5000E 0
A 1.B 2	8	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0
A 1.B 3	9	10.0000E -1	6.6667E -1	-2.2222E -1	-2.7778E -1	-2.2222E -1	8.3333E -1	-1.6667E -1
A 2.B 1	10	-4.0000E 0	-6.6667E -1	1.5556E 0	1.4444E 0	-4.4444E -1	-3.3333E -1	-3.3333E -1
A 2.B 2	11	4.0000E 0	-6.6667E -1	1.5556E 0	-5.5556E -1	1.5556E 0	-3.3333E -1	-3.3333E -1
A 2.B 3	12	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0
MEAN	13	1.1000E 1	3.3333E -1	2.2222E -1	2.7778E -1	2.2222E -1	1.6667E -1	1.6667E -1
		1	2	3	4	5	6	7

A 1.B 2	8	0.0000E 0					
A 1.B 3	9	0.0000E 0	9.4444E -1				
A 2.B 1	10	0.0000E 0	-1.1111E -1	1.7778E 0			
A 2.B 2	11	0.0000E 0	-1.1111E -1	-2.2222E -1	1.7778E 0		
A 2.B 3	12	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	
MEAN	13	0.0000E 0	5.5556E -2	1.1111E -1	1.1111E -1	0.0000E 0	1.8000E 1
		8	9	10	11	12	13

***** CORRELATION MATRIX *****

DF = 16

Y	1	1.0000								
A 1	2	-0.2554	1.0000							
A 2	3	-0.0000	-0.3780	1.0000						
B 1	4	-0.3136	0.3508	0.2652	1.0000					
B 2	5	-0.0965	-0.3780	0.3571	-0.3315	1.0000				
B 3	6	0.4845	-0.0000	-0.2390	-0.2774	-0.2390	1.0000			
A 1.B 1	7	-0.1615	0.6325	-0.2390	0.7211	-0.2390	-0.2000	1.0000		
A 1.B 2	8	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
A 1.B 3	9	0.0876	0.3430	-0.1296	-0.1504	-0.1296	0.5423	-0.1085	0.0000	1.0000
A 2.B 1	10	-0.2554	-0.2500	0.6614	0.5701	-0.1890	-0.1581	-0.1581	0.0000	-0.0857
A 2.B 2	11	0.2554	-0.2500	0.6614	-0.2193	0.6614	-0.1581	-0.1581	0.0000	-0.0857
A 2.B 3	12	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MEAN	13	0.2207	0.0393	0.0297	0.0345	0.0297	0.0248	0.0248	0.0000	0.0135
		1	2	3	4	5	6	7	8	9

A 2.B 1	10	1.0000			
A 2.B 2	11	-0.1250	1.0000		
A 2.B 3	12	0.0000	0.0000	0.0000	
MEAN	13	0.0196	0.0196	0.0000	1.0000
		10	11	12	13

60

FIT CONSTANT ADD A,B,A.B PARTITIONING SS

105.. 61

*** LINEAR DEPENDENCE DETECTED WHILE FITTING TERM A.B
4 PARAMETER(S) OF THIS TERM ARE ALIASED.

Indicates not all interactions estimable.

***** REGRESSION ANALYSIS *****

*** INVERSE MATRIX ***

		1	2	3	4	5	6
CONSTANT	1	2.5000E -1					
A 1	2	-2.5000E -1	7.5000E -1				
A 2	3	-4.2327E-16	8.1185E-16	10.0000E -1			
B 1	4	-2.5000E -1	2.5000E -1	-10.0000E -1	1.7500E 0		
B 2	5	-2.5000E -1	2.5000E -1	-5.0000E -1	7.5000E -1	7.5000E -1	
B 3	6	-2.5000E -1	2.5000E -1	4.7878E-16	2.5000E -1	2.5000E -1	7.5000E -1
A 1.B 1	7	2.5000E -1	-7.5000E -1	10.0000E -1	-1.7500E 0	-7.5000E -1	-2.5000E -1
A 1.B 2	8	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0
A 1.B 3	9	2.5000E -1	-7.5000E -1	-8.1185E-16	-2.5000E -1	-2.5000E -1	-7.5000E -1
A 2.B 1	10	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0
A 2.B 2	11	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0
A 2.B 3	12	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0

		7	8	9	10	11	12
A 1.B 1	7	2.5833E 0					
A 1.B 2	8	0.0000E 0	0.0000E 0				
A 1.B 3	9	7.5000E -1	0.0000E 0	2.2500E 0			
A 2.B 1	10	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0		
A 2.B 2	11	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	
A 2.B 3	12	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0

*** VARIANCE-COVARIANCE MATRIX ***

Y-VARIATE: Y

		1	2	3	4	5	6
CONSTANT	1	1.4000E 0					
A 1	2	-1.4000E 0	4.2000E 0				
A 2	3	-2.3703E-15	4.5464E-15	5.6000E 0			
B 1	4	-1.4000E 0	1.4000E 0	-5.6000E 0	9.8000E 0		
B 2	5	-1.4000E 0	1.4000E 0	-2.8000E 0	4.2000E 0	4.2000E 0	
B 3	6	-1.4000E 0	1.4000E 0	2.6812E-15	1.4000E 0	1.4000E 0	4.2000E 0
A 1.B 1	7	1.4000E 0	-4.2000E 0	5.6000E 0	-9.8000E 0	-4.2000E 0	-1.4000E 0
A 1.B 2	8	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0
A 1.B 3	9	1.4000E 0	-4.2000E 0	-4.5464E-15	-1.4000E 0	-1.4000E 0	-4.2000E 0
A 2.B 1	10	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0
A 2.B 2	11	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0
A 2.B 3	12	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0

		7	8	9	10	11	12
A 1.B 1	7	1.4467E 1					
A 1.B 2	8	0.0000E 0	0.0000E 0				
A 1.B 3	9	4.2000E 0	0.0000E 0	1.2600E 1			
A 2.B 1	10	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0		
A 2.B 2	11	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	
A 2.B 3	12	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0

*** REGRESSION COEFFICIENTS ***

Y-VARIATE: Y

	ESTIMATE	S.E.	T
CONSTANT	1.2000E 1	1.1832E 0	10.14
A 1	-3.0000E 0	2.0494E 0	-1.46
A 2	5.0000E 0	2.3664E 0	2.11
B 1	-8.0000E 0	3.1305E 0	-2.56
B 2	-4.0000E 0	2.0494E 0	-1.95
B 3	3.0000E 0	2.0494E 0	1.46
A 1.B 1	9.0000E 0	3.8035E 0	2.37
A 1.B 2	0.0000E 0	0.0000E 0	*
A 1.B 3	-1.5488E-14	3.5496E 0	-0.00
A 2.B 1	0.0000E 0	0.0000E 0	*
A 2.B 2	0.0000E 0	0.0000E 0	*
A 2.B 3	0.0000E 0	0.0000E 0	*

$$\underset{\sim}{b}^{o'} = [12 \quad \underbrace{-3 \quad 5 \quad 0}_{\alpha's} \quad \underbrace{-8 \quad -4 \quad 3 \quad 0}_{\beta's} \quad \underbrace{9 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0]}_{\gamma's}$$

*** ANALYSIS OF VARIANCE ***

	DF	SS		MS
REGRESSN	7	82.00	SSR_m	11.714
RESIDUAL	10	56.00	SSE	5.600
TOTAL	17	138.00	SST_m	8.118
CHANGE	-7	-82.00		11.714

% VARIANCE ACCOUNTED FOR 31.0

*** OBSERVED AND FITTED VALUES ***

	OBSERVED	FITTED	RESIDUAL
1	8.00	10.00	-2.00
2	13.00	10.00	3.00
3	9.00	10.00	-1.00
4	12.00	12.00	0.00
5	7.00	9.00	-2.00
6	11.00	9.00	2.00
7	6.00	9.00	-3.00
8	12.00	9.00	3.00
9	12.00	13.00	-1.00
10	14.00	13.00	1.00
11	9.00	8.00	1.00
12	7.00	8.00	-1.00
13	14.00	15.00	-1.00
14	16.00	15.00	1.00
15	10.00	12.00	-2.00
16	14.00	12.00	2.00
17	11.00	12.00	-1.00
18	13.00	12.00	1.00

*** SET OF FITTED TERMS ***

CONSTANT
A
B
A.B

```
*** ANALYSIS OF DEVIANCE ***

Y-VARIATE: Y

 TERMS                          RESIDUAL        CHANGE              MEAN    MN DEV.
                               DF   DEVIANCE  DF    DEVIANCE       CHANGE   RATIO
INITIAL MODEL
 CONSTANT                      17    138.000   *          *

MODIFICATIONS TO MODEL
+A                             15    127.500   2      10.500        5.250    0.94
+B                             12     90.714   3      36.786       12.262    2.19
+A.B                           10     56.000   2      34.714       17.357    3.10
** DENOMINATOR OF RATIO IS (RES. DEV./RES. DF) FROM LINE ABOVE, =       5.600
FIT CONSTANT ADD B,A,A.B PARTITIONING SS

110.......................................................................................................

*** LINEAR DEPENDENCE DETECTED WHILE FITTING TERM A.B
      4 PARAMETER(S) OF THIS TERM ARE ALIASED.


*** ANALYSIS OF DEVIANCE ***

Y-VARIATE: Y

 TERMS                          RESIDUAL        CHANGE              MEAN    MN DEV.
                               DF   DEVIANCE  DF    DEVIANCE       CHANGE   RATIO
INITIAL MODEL
 CONSTANT                      17    138.000   *          *

MODIFICATIONS TO MODEL
+B                             14    100.200   3      37.800       12.600    2.25
+A                             12     90.714   2       9.486        4.743    0.85
+A.B                           10     56.000   2      34.714       17.357    3.10
** DENOMINATOR OF RATIO IS (RES. DEV./RES. DF) FROM LINE ABOVE, =       5.600

 114  'CLOSE'
 ******* END OF         SET_5     AT LINE  100     USED 812     LEFT 7188
```

Handwritten annotations beside the first table:

$R(\alpha|\mu)$

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

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

Handwritten annotations beside the second table:

$R(\beta|\mu)$

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

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

```
GENSTAT V     MARK 3.09
(C)1976 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STATION)

'REFERENCE' SET_6
'CAPTION'
            ''
                 DATA SET 6
                 UNBALANCED DATA, 1 WAY CLASSIFICATION
                 WITH 1 COVARIATE
                 LINEAR MODELS BY S R SEARLE PAGE 353
            ''
'UNITS'          $ 7
'FACTOR'         A $ 3
'READ/P'         A,Y,X
'PRINT/P'        X,A,Y $ 10.0
'PAGE'
'TERMS/PRINT = SC'  A + X + Y
'Y'               Y
'PAGE'
'CAPTION'
                 ''FACTOR THEN COVARIATE  TABLE 8.6A''
'FIT/PRINT = Z, ANDEV=I'
'ADD/ANDEV=PT,PRINT=XACIVU'   A, X
'DROP/PRINT=Z'       A, X
'PAGE'
'CAPTION'
                 ''COVARIATE THEN FACTOR  TABLE 8.6B''
'FIT/PRINT = Z, ANDEV=I'
'ADD/ANDEV=PT,PRINT=Z'  X, A
'PAGE'
'TERMS/DVSET=F'      A*X + Y
'Y'               Y
'CAPTION'
                 '' SEPARATE SLOPES AND INTERCEPTS FOR EACH FACTOR LEVEL''
'FIT/PRINT = Z, ANDEV=I'
'ADD/PRINT=XACIVU,ANDEV=PT'  A + A.X
'DROP/PRINT=Z'       A + A.X
'PAGE'
'CAPTION'
                 '' TEST FOR DIFFERENT SLOPES''
'FIT/PRINT = Z, ANDEV=I'
'ADD/PRINT=XACIVU,ANDEV=PT'  A + X + A.X
'RUN'
```

4 sequences of models fitted

```
        DATA SET 6
        UNBALANCED DATA, 1 WAY CLASSIFICATION
        WITH 1 COVARIATE
        LINEAR MODELS BY S R SEARLE PAGE 353

  Y  0  MNMINMAX       79.0000       68.0000       93.0000     7  VALUES    0 MISSING
  X  0  MNMINMAX        3.5714        2.0000        6.0000     7  VALUES    0 MISSING


     X          A          Y
     3          1         74
     4          1         68
     2          1         77
     2          2         76
     4          2         80
     4          3         85
     6          3         93
```

***** SUMS OF SQUARES AND PRODUCTS MATRIX *****

DF = 6

X	1	1.1714E 1				
Y	2	4.3000E 1	3.9200E 2			
A 1	3	-1.7143E 0	-1.8000E 1	1.7143E 0		
A 2	4	-1.1429E 0	-2.0000E 0	-8.5714E -1	1.4286E 0	
MEAN	5	3.5714E 0	7.9000E 1	4.2857E -1	2.8571E -1	7.0000E 0
		1	2	3	4	5

***** CORRELATION MATRIX *****

DF = 5

X	1	1.0000				
Y	2	0.6346	1.0000			
A 1	3	-0.3825	-0.6944	1.0000		
A 2	4	-0.2794	-0.0845	-0.5477	1.0000	
MEAN	5	0.3944	1.5081	0.1237	0.0904	1.0000
		1	2	3	4	5

FACTOR THEN COVARIATE TABLE 8.6A

Single Regression Model: $E(y_{ij}) = \mu + \alpha_i + bx_{ij}$

20..

***** REGRESSION ANALYSIS *****

*** INVERSE MATRIX ***

CONSTANT	1	4.6667E 0			
A 1	2	-2.1667E 0	1.5000E 0		
A 2	3	-2.1667E 0	1.1667E 0	1.6667E 0	
X	4	-8.3333E -1	3.3333E -1	3.3333E -1	1.6667E -1
		1	2	3	4

*** VARIANCE-COVARIANCE MATRIX ***

Y-VARIATE: Y

CONSTANT	1	1.2522E 2			
A 1	2	-5.8139E 1	4.0250E 1		
A 2	3	-5.8139E 1	3.1306E 1	4.4722E 1	
X	4	-2.2361E 1	8.9444E 0	8.9444E 0	4.4722E 0
		1	2	3	4

*** REGRESSION COEFFICIENTS ***

Y-VARIATE: Y

	ESTIMATE	S.E.	T
CONSTANT	86.49998	11.19027	7.73
A 1	-15.00000	6.34429	-2.36
A 2	-10.00000	6.68746	-1.50
X	0.50000	2.11476	0.24

$\underset{\sim}{b}^{o\prime} = [\mu^o \quad \alpha_1^o \quad \alpha_2^o \quad \alpha_3^o \quad \hat{b}]$

$= [86\tfrac{1}{2} \quad -15 \quad -10 \quad 0 \quad \tfrac{1}{2}]$

i.e., $\underset{\sim}{a}^{o\prime} = [86\tfrac{1}{2} \quad -15 \quad -10 \quad 0]$, equivalent to (44) in LM, p. 354,

and $\hat{b} = \tfrac{1}{2}$, of (42) in LM, p. 353.

b is estimable, so the SE and t-test are appropriate.

*** ANALYSIS OF VARIANCE ***

	DF	SS	MS
REGRESSN	3	311.50	103.83
RESIDUAL	3	80.50	26.83
TOTAL	6	392.00	65.33
CHANGE	-3	-311.50	103.83

% VARIANCE ACCOUNTED FOR 58.9

*** OBSERVED AND FITTED VALUES ***

	OBSERVED	FITTED	RESIDUAL
1	74.0	73.0	1.0
2	68.0	73.5	-5.5
3	77.0	72.5	4.5
4	76.0	77.5	-1.5
5	80.0	78.5	1.5
6	85.0	88.5	-3.5
7	93.0	89.5	3.5

This is $\hat{\underset{\sim}{y}} = \underset{\approx}{X}\underset{\sim}{a}^o + \underset{\sim}{x}\hat{b}$

e.g., $\hat{y}_{11} = \mu^o + \alpha_1^o + \hat{b}x_{11} = 86\tfrac{1}{2} - 15 + \tfrac{1}{2}(3) = 73$

Sum of squares = 80.5 = SSE

*** SET OF FITTED TERMS ***

```
CONSTANT
A
X
```

```
*** ANALYSIS OF DEVIANCE ***

Y-VARIATE: Y

 TERMS                                   RESIDUAL        CHANGE            MEAN    MN DEV.
                                        DF    DEVIANCE  DF    DEVIANCE     CHANGE   RATIO
INITIAL MODEL
 CONSTANT                                6      392.00   *            *

MODIFICATIONS TO MODEL
+A                                       4       82.00   2      310.00}    155.00    7.56
** DENOMINATOR OF RATIO IS (RES. DEV./RES. DF) FROM LINE ABOVE, =         20.50
+X                                       3       80.50   1        1.50}      1.50    0.07
                                                    SSE                 R(α|μ)
                                                                        R(b|μ,α)
```

Table 8.6a, LM p. 35-.

COVARIATE THEN FACTOR TABLE 8.6B

```
*** ANALYSIS OF DEVIANCE ***

Y-VARIATE: Y

 TERMS                                   RESIDUAL        CHANGE            MEAN    MN DEV.
                                        DF    DEVIANCE  DF    DEVIANCE     CHANGE   RATIO
INITIAL MODEL
 CONSTANT                                6      392.00   *            *

MODIFICATIONS TO MODEL
+X                                       5      234.16   1      157.84}    157.84    5.88
+A                                       3       80.50   2      153.66}     76.83    2.86
** DENOMINATOR OF RATIO IS (RES. DEV./RES. DF) FROM LINE ABOVE, =         26.83
                                                    SSE                 R(b|μ)
                                                                        R(α|μ,b)
```

Table 8.6b, LM p. 355.

SEPARATE SLOPES AND INTERCEPTS FOR EACH FACTOR LEVEL

Regression for each group: $E(y_{ij}) = \mu + \alpha_i + b_i x_{ij}$

33..

***** REGRESSION ANALYSIS *****

*** INVERSE MATRIX ***

		1	2	3	4	5	6
CONSTANT	1	1.3000E 1					
A 1	2	-1.3000E 1	1.7833E 1				
A 2	3	-1.3000E 1	1.3000E 1	1.8000E 1			
A 3	4	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0		
X.A 1	5	2.9421E-15	-1.5000E 0	-4.2188E-15	0.0000E 0	5.0000E -1	
X.A 2	6	3.4417E-15	-4.7740E-15	-1.5000E 0	0.0000E 0	4.4409E-16	5.0000E -1
X.A 3	7	-2.5000E 0	2.5000E 0	2.5000E 0	0.0000E 0	-5.5511E-16	-6.6613E-16

		7
X.A 3	7	5.0000E -1

*** VARIANCE-COVARIANCE MATRIX ***

Y-VARIATE: Y

		1	2	3	4	5	6
CONSTANT	1	1.9500E 1					
A 1	2	-1.9500E 1	2.6750E 1				
A 2	3	-1.9500E 1	1.9500E 1	2.7000E 1			
A 3	4	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0		
X.A 1	5	4.4131E-15	-2.2500E 0	-6.3283E-15	0.0000E 0	7.5000E -1	
X.A 2	6	5.1625E-15	-7.1609E-15	-2.2500E 0	0.0000E 0	6.6613E-16	7.5000E -1
X.A 3	7	-3.7500E 0	3.7500E 0	3.7500E 0	0.0000E 0	-8.3267E-16	-9.9920E-16

		7
X.A 3	7	7.5000E -1

*** REGRESSION COEFFICIENTS ***

Y-VARIATE: Y

	ESTIMATE	S.E.	T
CONSTANT	69.0000	4.4159	15.63
A 1	17.5000	5.1720	3.38
A 2	3.0000	5.1962	0.58
A 3	0.0000	0.0000	*
X.A 1	-4.5000	0.8660	-5.20
X.A 2	2.0000	0.8660	2.31
X.A 3	4.0000	0.8660	4.62

$\underset{\sim}{a}^{o\prime} = [69 \quad 17\tfrac{1}{2} \quad 3 \quad 0]$

$\hat{b}' = [\hat{b}_1 \quad \hat{b}_2 \quad \hat{b}_3] = [-4\tfrac{1}{2} \quad 2 \quad 4]$. See LM p. 358.

*** ANALYSIS OF VARIANCE ***

	DF	SS		MS
REGRESSN	5	390.500	SSR_m	78.100
RESIDUAL	1	1.500	SSE	1.500
TOTAL	6	392.000	SST_m	65.333
CHANGE	-5	-390.500		78.100

% VARIANCE ACCOUNTED FOR 97.7

*** OBSERVED AND FITTED VALUES ***

$\hat{\underset{\sim}{y}} = \underset{\approx}{X}\underset{\sim}{\alpha}^{o} + \underset{\approx}{Z}\hat{\underset{\sim}{b}}$

e.g., $\hat{y}_{11} = \mu^{o} + \alpha_1^{o} + \hat{b}_1 x_{11} = 69 + 17\frac{1}{2} - 4\frac{1}{2}(3) = 73$

	OBSERVED	FITTED	RESIDUAL
1	74.00	73.00	1.00
2	68.00	68.50	-0.50
3	77.00	77.50	-0.50
4	76.00	76.00	0.00
5	80.00	80.00	0.00
6	85.00	85.00	0.00
7	93.00	93.00	0.00

*** SET OF FITTED TERMS ***

CONSTANT
A
X.A

*** ANALYSIS OF DEVIANCE ***

Y-VARIATE: Y

TERMS	RESIDUAL DF	RESIDUAL DEVIANCE	CHANGE DF	CHANGE DEVIANCE	MEAN CHANGE	MN DEV. RATIO
INITIAL MODEL						
CONSTANT	6	392.000	*	*		
MODIFICATIONS TO MODEL						
+A	4	82.000	2	310.000	155.000	103.33
+X.A	1	1.500	3	80.500	26.833	17.89

** DENOMINATOR OF RATIO IS (RES. DEV./RES. DF) FROM LINE ABOVE, = 1.500

SSE

$R(\alpha|\mu)$

$R(\underset{\sim}{b}|\mu,\alpha)$

TEST FOR DIFFERENT SLOPES

39.

***** REGRESSION ANALYSIS *****

*** INVERSE MATRIX ***

		1	2	3	4	5	6
CONSTANT	1	1.3000E 1					
A 1	2	-1.3000E 1	1.7833E 1				
A 2	3	-1.3000E 1	1.3000E 1	1.8000E 1			
A 3	4	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0		
X	5	-2.5000E 0	2.5000E 0	2.5000E 0	0.0000E 0	5.0000E -1	
X.A 1	6	2.5000E 0	-4.0000E 0	-2.5000E 0	0.0000E 0	-5.0000E -1	1.0000E 0
X.A 2	7	2.5000E 0	-2.5000E 0	-4.0000E 0	0.0000E 0	-5.0000E -1	5.0000E -1
X.A 3	8	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0

		7	8
X.A 2	7	1.0000E 0	
X.A 3	8	0.0000E 0	0.0000E 0

*** VARIANCE-COVARIANCE MATRIX ***

Y-VARIATE: Y

		1	2	3	4	5	6
CONSTANT	1	1.9500E 1					
A 1	2	-1.9500E 1	2.6750E 1				
A 2	3	-1.9500E 1	1.9500E 1	2.7000E 1			
A 3	4	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0		
X	5	-3.7500E 0	3.7500E 0	3.7500E 0	0.0000E 0	7.5000E -1	
X.A 1	6	3.7500E 0	-6.0000E 0	-3.7500E 0	0.0000E 0	-7.5000E -1	1.5000E 0
X.A 2	7	3.7500E 0	-3.7500E 0	-6.0000E 0	0.0000E 0	-7.5000E -1	7.5000E -1
X.A 3	8	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0	0.0000E 0

		7	8
X.A 2	7	1.5000E 0	
X.A 3	8	0.0000E 0	0.0000E 0

*** REGRESSION COEFFICIENTS ***

Y-VARIATE: Y

	ESTIMATE	S.E.	T
CONSTANT	69.0000	4.4159	15.63
A 1	17.5000	5.1720	3.38
A 2	3.0000	5.1961	0.58
A 3	0.0000	0.0000	*
X	4.0000	0.8660	4.62
X.A 1	-8.5000	1.2247	-6.94
X.A 2	-2.0000	1.2247	-1.63
X.A 3	0.0000	0.0000	*

70

The slopes are a common slope of 4 and the 3 different slopes given as deviations from the common slope, with

$$\underset{\sim}{b}^{o\prime} = [b^o \;\; b_1^o \;\; b_2^o \;\; b_3^o]$$

being now solutions and not estimates because of over-parameterization of the regression; i.e., for the previous model with separate slopes

$$\hat{b}_i = b^o + b_i^o$$

e.g., $-4\frac{1}{2} = 4 - 8.5$.

```
*** ANALYSIS OF VARIANCE ***

                 DF          SS               MS

  REGRESSN        5     390.500           78.100
  RESIDUAL        1       1.500            1.500
     TOTAL        6     392.000           65.333

    CHANGE       -5    -390.500           78.100

     % VARIANCE ACCOUNTED FOR 97.7

*** OBSERVED AND FITTED VALUES ***
                 OBSERVED        FITTED      RESIDUAL

            1       74.00         73.00          1.00
            2       68.00         68.50         -0.50
            3       77.00         77.50         -0.50
            4       76.00         76.00          0.00
            5       80.00         80.00          0.00
            6       85.00         85.00          0.00
            7       93.00         93.00         -0.00

*** SET OF FITTED TERMS ***

CONSTANT
A
X
X.A


*** ANALYSIS OF DEVIANCE ***

Y-VARIATE: Y

 TERMS                        RESIDUAL       CHANGE            MEAN   MN DEV.
                             DF   DEVIANCE  DF    DEVIANCE     CHANGE   RATIO
INITIAL MODEL
 CONSTANT                    6    392.000    *           *

MODIFICATIONS TO MODEL
+A                           4     82.000   2     310.000    155.000  103.33
+X                           3     80.500   1       1.500      1.500    1.00
+X.A                         1      1.500   2      79.000     39.500   26.33
** DENOMINATOR OF RATIO IS (RES. DEV./RES. DF) FROM LINE ABOVE, =      1.500

  43  'CLOSE'
 ******* END OF        SET_6    AT LINE    28     USED 809     LEFT 7191
```

Table 8.8, LM p. 359.

SSE

$R(\alpha|\mu)$

$R(b|\mu,\alpha)$

$R(\underset{\sim}{b}|\mu,\alpha,b) = R(\underset{\sim}{b}|\mu,\alpha) - R(b|\mu,\alpha)$

```
GENSTAT V    MARK 3.09
(C)1976 LAWES AGRICULTURAL TRUST (ROTHAMSTED EXPERIMENTAL STATION)

'REFERENCE' SET_7
'CAPTION'
                ''
                    DATA SET 7
                    UNBALANCED DATA, TWO WAY CROSSED CLASSIFICATION
                    WITH 1 COVARIATE
                       LINEAR MODELS BY S R SEARLE PAGE 375
                ''
'UNITS'         $ 48
'READ/S' Y,X
'FACTOR'        A $ 3
  :             B $ 4
'GENERATE'      A,B,4
'PRINT/P'       A,B,X,Y  $10.0
'TERMS/DVSET=F'      A*B + A*X +Y
'Y'             Y
'FIT/PRINT = Z, ANDEV=I'
'ADD/ANDEV=PT'       X + A*B
'DROP/PRINT=Z'       X + A*B
'FIT/PRINT = Z, ANDEV=I'
                ''SEPARATE SLOPES FOR EACH ROW''
'ADD/ANDEV=PT'       A.X + A.B
'DROP/PRINT=Z'       A.X + A.B
'FIT/PRINT = Z, ANDEV=I'
'ADD/PRINT=Z,ANDEV=PT' A.B + A.X
'DROP/PRINT=Z'       A.X + A.B
                ''TEST SEPARATE SLOPES''
'FIT/PRINT = Z, ANDEV=I'
'ADD/PRINT=Z,ANDEV=PT'  A.B + X + A.X
'DROP/PRINT=Z'          A.B + X + A.X
'FIT/PRINT = Z, ANDEV=I'
'ADD/PRINT=Z,ANDEV=PT'  X + A.X + A.B
'RUN'
```

DATA SET 7
UNBALANCED DATA, TWO WAY CROSSED CLASSIFICATION
WITH 1 COVARIATE
LINEAR MODELS BY S R SEARLE PAGE 375

Y	0	MNMINMAX	11.0000	6.0000	16.0000	48	VALUES	30 MISSING
X	0	MNMINMAX	4.7778	2.0000	8.0000	48	VALUES	30 MISSING

A	B	X	Y
1	1	2	8
1	1	4	13
1	1	3	9
1	1	*	*
1	2	*	*
1	2	*	*
1	2	*	*
1	2	*	*
1	3	7	12
1	3	*	*
1	3	*	*
1	3	*	*
1	4	3	7
1	4	5	11
1	4	*	*
1	4	*	*
2	1	5	6
2	1	3	12
2	1	*	*
2	1	*	*
2	2	6	12
2	2	4	14
2	2	*	*
2	2	*	*
2	3	*	*
2	3	*	*
2	3	*	*
2	3	*	*
2	4	*	*
2	4	*	*
2	4	*	*
2	4	*	*
3	1	*	*
3	1	*	*
3	1	*	*
3	1	*	*
3	2	6	9
3	2	2	7
3	2	*	*
3	2	*	*
3	3	6	14
3	3	8	16
3	3	*	*
3	3	*	*
3	4	4	10
3	4	6	14
3	4	5	11
3	4	7	13

The regression program operates either on data that contain the "missing value indicators", as here, or on data that exclude them, as in cases 3 to 6.

***** REGRESSION ANALYSIS *****

Model: $E(y_{ijk}) = \mu + \alpha_i + \beta_j + \gamma_{ij} + bx_{ijk}$

*** REGRESSION COEFFICIENTS ***

$\hat{b} = .56522$

Y-VARIATE: Y

	ESTIMATE	S.E.	T
CONSTANT	8.89130	2.90882	3.06
X	$\boxed{0.56522}$	0.48480	1.17
A 1	-2.15217	2.14083	-1.01
A 2	4.43478	2.37504	1.87
A 3	0.00000	0.00000	*
B 1	-6.58696	3.30592	-1.99
B 2	-3.15217	2.14083	-1.47
B 3	2.15217	2.14083	1.01
B 4	0.00000	0.00000	*
A 1.B 1	8.15217	3.80707	2.14
A 1.B 2	0.00000	0.00000	*
A 1.B 3	-0.84783	3.56256	-0.24
A 1.B 4	0.00000	0.00000	*
A 2.B 1	0.00000	0.00000	*
A 2.B 2	0.00000	0.00000	*
A 2.B 3	0.00000	0.00000	*
A 2.B 4	0.00000	0.00000	*
A 3.B 1	0.00000	0.00000	*
A 3.B 2	0.00000	0.00000	*
A 3.B 3	0.00000	0.00000	*
A 3.B 4	0.00000	0.00000	*

*** ANALYSIS OF VARIANCE ***

	DF	SS		MS
REGRESSN	8	89.35	$R(b,\alpha,\beta,\gamma \mid \mu)$	11.168
RESIDUAL	9	48.65	SSE	5.406
TOTAL	17	138.00	SST_m	8.118
CHANGE	-8	-89.35		11.168

% VARIANCE ACCOUNTED FOR 33.4

*** ANALYSIS OF DEVIANCE ***

Y-VARIATE: Y

TERMS	RESIDUAL DF	RESIDUAL DEVIANCE	CHANGE DF	CHANGE DEVIANCE	MEAN CHANGE	MN DEV. RATIO
INITIAL MODEL						
CONSTANT	17	138.000	*	*		
MODIFICATIONS TO MODEL						
+X	16	85.111	1	52.889	52.889	9.94
** DENOMINATOR OF RATIO IS (RES. DEV./RES. DF) FROM LINE ABOVE, = 5.319						
+A	14	84.478	2	0.633	0.317	0.06
+B	11	79.082	3	5.396	1.799	0.34
+A.B	9	48.652	2	30.430	15.215	2.86

$R(b \mid \mu)$

$R(\alpha \mid \mu, b)$

$R(\beta \mid \mu, \alpha, b)$

$R(\gamma \mid \mu, \alpha, \beta, b)$

74

***** REGRESSION ANALYSIS *****

Model: $E(y_{ijk}) = \alpha_i + \beta_j + \gamma_{ij} + b_i x_{ijk}$

*** REGRESSION COEFFICIENTS ***

$\hat{b}_1$
$\hat{b}_2$
$\hat{b}_3$

Y-VARIATE: Y

	ESTIMATE	S.E.	T
CONSTANT	7.60000	1.81292	4.19
X.A 1	2.25000	0.60208	3.74
X.A 2	-2.00000	0.60208	-3.32
X.A 3	0.80000	0.31091	2.57
A 1.B 1	-4.35000	2.65189	-1.64
A 1.B 2	0.00000	0.00000	*
A 1.B 3	-11.35000	4.74333	-2.39
A 1.B 4	-7.60000	3.13236	-2.43
A 2.B 1	9.40000	3.13236	3.00
A 2.B 2	15.40000	3.61582	4.26
A 2.B 3	0.00000	0.00000	*
A 2.B 4	0.00000	0.00000	*
A 3.B 1	0.00000	0.00000	*
A 3.B 2	-2.80000	1.14237	-2.45
A 3.B 3	1.80000	1.14237	1.58
A 3.B 4	0.00000	0.00000	*

*** ANALYSIS OF VARIANCE ***

	DF	SS	MS
REGRESSN	10	127.85	12.785
RESIDUAL	7	10.15	1.450
TOTAL	17	138.00	8.118
CHANGE	-10	-127.85	12.785

% VARIANCE ACCOUNTED FOR 82.1

*** ANALYSIS OF DEVIANCE ***

Factors after covariates.

Y-VARIATE: Y

TERMS		RESIDUAL DF	RESIDUAL DEVIANCE	CHANGE DF	CHANGE DEVIANCE	MEAN CHANGE	MN DEV. RATIO	
INITIAL MODEL								
CONSTANT		17	138.000	*	*			
MODIFICATIONS TO MODEL								
+X.A		14	84.034	3	53.966	17.989	12.41	$R(\underset{\sim}{b} \mid \mu)$
+A.B	SSE	7	10.150	7	73.884	10.555	7.28	$R(\mu,\alpha,\beta,\gamma \mid \mu,\underset{\sim}{b})$

** DENOMINATOR OF RATIO IS (RES. DEV./RES. DF) FROM LINE ABOVE, = 1.450

75

*** ANALYSIS OF DEVIANCE ***

Factors before covariates.

Y-VARIATE: Y

TERMS	RESIDUAL DF	RESIDUAL DEVIANCE	CHANGE DF	CHANGE DEVIANCE	MEAN CHANGE	MN DEV. RATIO	
INITIAL MODEL							
CONSTANT	17	138.000	*	*			
MODIFICATIONS TO MODEL							
+A.B	10	56.000	7	82.000	11.714	8.08	$R(\alpha,\beta,\gamma\mid\mu)$
+X.A	SSE 7	10.150	3	45.850	15.283	10.54	$R(\underset{\sim}{b}\mid\mu,\alpha,\beta,\gamma)$

** DENOMINATOR OF RATIO IS (RES. DEV./RES. DF) FROM LINE ABOVE, = 1.450

*** ANALYSIS OF DEVIANCE ***

Testing separate slopes.

Y-VARIATE: Y

TERMS	RESIDUAL DF	RESIDUAL DEVIANCE	CHANGE DF	CHANGE DEVIANCE	MEAN CHANGE	MN DEV. RATIO	
INITIAL MODEL							
CONSTANT	17	138.000	*	*			
MODIFICATIONS TO MODEL							
+A.B	10	56.000	7	82.000	11.714	8.08	$R(\alpha,\beta,\gamma\mid\mu)$
+X	9	48.652	1	7.348	7.348	5.07	$R(b\mid\mu,\alpha,\beta,\gamma)$
+X.A	SSE 7	10.150	2	38.502	19.251	13.28	$R(\underset{\sim}{b}\mid\mu,\alpha,\beta,\gamma,b)$

** DENOMINATOR OF RATIO IS (RES. DEV./RES. DF) FROM LINE ABOVE, = 1.450

*** ANALYSIS OF DEVIANCE ***

Y-VARIATE: Y

TERMS	RESIDUAL DF	RESIDUAL DEVIANCE	CHANGE DF	CHANGE DEVIANCE	MEAN CHANGE	MN DEV. RATIO	
INITIAL MODEL							
CONSTANT	17	138.000	*	*			
MODIFICATIONS TO MODEL							$R(b\mid\mu)$
+X	16	85.111	1	52.889	52.889	36.48	$R(\underset{\sim}{b}\mid\mu,b)$
+X.A	14	84.034	2	1.077	0.538	0.37	
+A.B	SSE 7	10.150	7	73.884	10.555	7.28	$R(\alpha,\beta,\gamma\mid\mu,b,\underset{\sim}{b})$

** DENOMINATOR OF RATIO IS (RES. DEV./RES. DF) FROM LINE ABOVE, = 1.450

```
 206   'CLOSE'
 ******* END OF          SET_7    AT LINE  179    USED 1420    LEFT 6580
```