The Philippine Statistician
Vol. xxii, Nos. 3 - 4, 1973

A COMPUTATIONAL TECHNIQUE FOR THE ANALYSIS OF TWO-WAY CLASSIFICATION WITH DISPROPORTIONATE SUBCLASS FREQUENCIES

By

HENRY F. MAGALIT*

Introduction

The analysis of a two-way classification with inter-action is sometimes not possible even with the use of computers because of the size of the matrix involved. The purpose of this paper is to partition the matrix and work with submatrices of manageable sizes that most often the calculations can be done with a desk calculator. The method can also be used using a computer when the ordinary least squares analysis cannot be used.

The method evolved in this paper uses the least squares procedure. It is an exact and convenient method for the randomized complete block design with missing values provided the rank of the design matrix *X is equal* to no. of *blocks* + *no. of treatments* — 1. It can also be used to analyze a factorial experiment in CRD with two factors and unequal number of replications. All balanced incomplete block designs not utilizing inter block information can also be analyzed using this method and tests for interaction is also possible with this method. Built-in checks are incorporated in the procedure so that at certain stages the computation can be checked.

Consider the model

$$(1)\ Y = X\beta + \epsilon$$

where Y is an n x 1 vector of observations

* Assistant Professor, University of the Philippines at Los Baños

ϵ is an n x 1 random vector

X is an n x (a + b) matrix of known fixed quantities

β is a (a + b) x 1 vector of unknown parameters

Assume that $\epsilon \sim N(0, \sigma^2 I)$ and the data is such that the rank of X is (a + b) − 1.

$$\text{Let } \beta = \begin{Bmatrix} \underset{a \times 1}{\alpha} \\ \underset{b \times 1}{\rho} \end{Bmatrix}$$

Data: ρ

		1	2	. . .	b	Total
	1	Y_{111} Y_{112} : $Y_{11}n_{11}$	Y_{121} Y_{122} : $Y_{12}n_{12}$	. . . .	Y_{1b1} Y_{1b2} : $Y_{1b}n_{1b}$	$Y_{1..}$
α	2	Y_{211} Y_{212} : $Y_{21}n_{21}$	. . . .			Y_2 . . . .
	a		Y_{a21} Y_{a22} : $Y_{a2}n_{a2}$		Y_{ab1} Y_{ab2} $Y_{abn_{ab}}$	$Y_{a..}$
	Total	Y.1.	Y.2.	. . .	Y.b.	

No. of Observations

P

		1	2	j	. b	Total
	1	n_{11}	n_{12}	n_{1j}	n_{1b}	$n_{1.}$
	2	n_{21}	n_{22}	n_{2j}	π	$n_{2.}$
a	i			n_{ij}		$n_{i.}$
	a	n_{a1}	n_{a2}	n_{aj}	n_{ab}	$n_{a..}$
Total		$n_{.1}$	$n_{.2}$	$n_{.j}$	n.b	n..

NOTATIONS

Let

$$A = \begin{Bmatrix} Y_{1..} \\ Y_{2..} \\ \vdots \\ Y_{a..} \end{Bmatrix} \qquad B = \begin{Bmatrix} Y_{.1.} \\ Y_{.2.} \\ \vdots \\ Y_{.b.} \end{Bmatrix}$$

$$\underset{a \times b}{N} = \{ n_{ij} \}$$

$$\underset{a \times a}{D_a} = \begin{Bmatrix} n_{2.} & 0 & \dots & 0 \\ 0 & n_{2.} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & \dots 0 & \dots & n_{a.} \end{Bmatrix}$$

$$\text{and} \quad \underset{b\,\times b}{D_b} = \left\{ \begin{array}{ccc} n_{1\cdot} & 0 \ldots & 0 \\ 0 & n_{\cdot 2} \ldots & 0 \\ \cdot & \cdot\cdot & \cdot \\ \cdot & \cdots & \cdot \\ \cdot\cdot & \cdot & \cdot\ \cdot \\ \cdot & \cdot & \cdot\ \cdot \\ 0 \ldots & 0 \ldots & n_{\cdot b} \end{array} \right\}$$

To get β (adjusted for α) form

$$(2) \qquad \begin{array}{c|c|c} D_a & N & A \\ \hline & D_b & B \end{array}$$

and to get α (adjusted for β) form

$$(3) \qquad \begin{array}{c|c|c} D_b & N' & B \\ \hline & D_a & A \end{array}$$

where N' is the transpose of N.
Let us consider (2). Now partition (2) into I and II as follows:

$$\begin{array}{c|c|cl} D_a & N & A & \} \ \text{I} \\ \hline & D_b & B & \} \ \text{II} \end{array}$$

"Doolittle" I and use this result to adjust II. This is really eliminating $\hat{\alpha}$ from the system of equations. The result will be a system of b equations in b unknowns, Denote this as (4): $C\rho = g$. This was derived by Kempthorne [1].

$$\text{where } \underset{b \times b}{C} = \{ c_{ij} \} \qquad \hat{\rho} = \begin{Bmatrix} b_1 \\ b_2 \\ \vdots \\ b_b \end{Bmatrix}$$

(5)

$$c_{jj} = n_{.j} - \sum_i^a \frac{n^2_{ij}}{n_{i.}} \quad \text{for } j = k$$

$$c_{jk} = - \sum_i^a \frac{n_{ij} n_{ik}}{n_i} \quad \text{for } j \neq k$$

$$\text{and } g_j = Y_{.j.} - \sum_i \frac{n_{ij}}{n_{i.}} Y_{i..} \qquad \text{for } j = 1, 2, \ldots, b$$

It can be shown that the rank of C is (b − 1), provided that the rank of X in (1) is (a + b) − 1. Therefore any row and column of C should add to zero and $\sum_j^b g_j$ is also zero.

Now add 1/b to each element of C. This is equivalent to imposing the restriction $\sum_j^b b_j = 0$. Denote the resulting system of equation (6): $D\hat{\rho} = g$. D is non-singular, and hence an inverse. "Doolittle" (6) to get ρ and also the sum of squares of the cross-product in Doolittle (ΣCPID) which is the sum of squares due to ρ adjusted for α. Subtracting 1/b from each element of D^{-1} we get a matrix denoted as $D' = \{d'_{ij}\}$. We will refer to D' later on.

[1]Kempthorne, 0. The Design and Analysis of Experiments John Wiley and Sons, Inc. 1952, p. 80.

We form the following AOVs:

Analysis of Variance
β(adj for α)

Sources of variation	d.f.	S.S.
Total	$n - 1$	$\sum_{i}^{n} \sum_{j}^{b} \sum_{k}^{n_{ij}} Y^2_{ijk} - CF$
α(unadj)	$a - 1$	$\sum_{i}^{a} Y^2_{i..} - CF$
β(adj for α)	$b - 1$	$\Sigma CPID$
$\alpha\beta$	$(a-1)(b-1)-e$	by subtraction
Within Cells	$n - ab + e$	$\sum_{i} \sum_{j} \sum_{k} Y^2_{ijk} - \sum \sum \frac{Y_{ij}^2}{n_{ij}}$

where e is the number of empty cells

$$CF \text{ if } \frac{(\Sigma\ \Sigma\ \Sigma\ Y_{ijk})^2}{n..}$$

and $n = n..$

Analysis of Variance
α(adj for β)

Sources of Variation	d.f.	S. S.
Total	$n - 1$	same as above
β(unadj)	$b - 1$	$\sum_{j} \frac{Y^2_{.j.}}{n_{.j}} - CF$
α(adj for β)	$a - 1$	α SS(unadj) + βSS(adj) — βSS (unadj)

$\alpha\beta$	$(-1)\ (b-1)\ -\ e$	Same as above
Within cells	$n-ab+e$	Same as above

If $n_{ij} < 2$ for all i and j the within cells cannot be obtained and hence will not appear in the analysis.

If there is no interaction and (adj) MS is significant the following hypothesis maybe tested using a t-test:

$$H_0:\ \rho_1 - \rho_2 = 0$$

Test Statistics: $$t = \frac{b_1 - b_2}{S_{(b1-b2)}}$$

where $S_{(b1-b2)} = \sqrt{EMS}\ (d'_{11} + d'_{22} - 2\ d'_{12})$

and d'_{ij} are elements of D'; EMS is error mean square

EXAMPLE 1 (Two-Way Classification without interaction)

Randomized complete block with missing observations:

Data:

ρ (Treatments)

α	1	2	3	
1	5	1	2	8
2	4	-	3	7
3	6	2	1	9
4	-	3	2	5
5	5	2	3	10
	20	8	11	39

No. of observation

α	1	2	3	
1	1	1	1	3
2	1	0	1	2
3	1	1	1	3
4	0	1	1	2
5	1	1	1	3
	4	4	5	13

Form D_a | N | A
| D_b | B :

3	0	0	0	0	1	1	1	8		I
	2	0	0	0	1	0	1	7		I
		3	0	0	1	1	1	9		I
			2	0	0	1	1	5		I
				3	1	1	1	10		I
					4	0	0	20		II
						4	0	8		II
							5	11		II

Doolittle I, we get:

3	0	0	0	0	1	1	1	8
1	0	0	0	0	1/3	1/3	1/3	8/3
	2	0	0	0	1	0	1	7
	1	0	0	0	1/2	0	1/2	7/2
		3	0	0	1	1	1	9
		1	0	0	1/3	1/3	1/3	9/3
			2	0	0	1	1	5
			1	0	0	1/2	1/2	5/2
				3	1	1	1	10
				1	1/3	1/3	1/3	10/3

Adjusting II using this result we get

$c_{11} = 4 - [1(1/3) + 1\ (1/2) + 1\ (1/3) + 0\ (0) =$

$$4 - \frac{2+3+2+2}{6} = 4 - \frac{9}{6} = \frac{15}{6}$$

$c_{12} = - [1(1/3) + 1(0) + 1(1/3) + 0(1/2) + 1(1/3)]$

$$= - \frac{6}{6}$$

$c_{13} = - [1(1/3) + 1(1/2) + 1(1/3) + 0(1/2) + 1(1/3)]$

$$= - \frac{2+3+2+2}{6} = - \frac{9}{6}$$

$c_{22} = 4 - [1(1/3) + 0(0) + 1(1/3) + 1(1/2 + 1(1/3)]$

$$= 4 - \frac{9}{6} = \frac{15}{6}$$

$c_{23} = - [1(1/3) + 0(1/2) + 1(1/3 + 1(1/2 + 1(1/3)]$

$$= - \frac{9}{6}$$

$c_{33} = 5 - [1(1/3) + 1(1/2) + 1(1/3) + 1(1/2) + 1(1/3)]$

$$= 5 - 2 = \frac{18}{6}$$

$g_1 = 20 - [1(8/3) + 1(7/2) + 1(9/3) + 0(5/2) + (10/3)]$

$$= 20 - 9 - \frac{7}{2} = \frac{15}{2}$$

$$g_2 = 8 - [1(8/3) + 0(7/2) + 1(9/3) + 1(5/2) + 1(10/3)]$$

$$= 8 - \frac{27}{3} - \frac{5}{2} = -\frac{7}{2}$$

$$g_3 = 11 - [1(8/3) + 1(7/2) + 1(9/3) + 1(5/2) + 1(10/3)]$$
$$= 11 - 9 - 6 = -4$$

$$\begin{array}{r|l} 15/6 - 6/6 - 9/6 & 15/2 \\ 15/6 - 9/6 & -7/2 \\ 18/6 & -8/2 \end{array}$$

add 1/3 to each element of the LHS of the above, convert to decimals and Doolittle we get:

2.83333333	-.66666667	-1.16666667	7.5	1	0	0
	2.83333333	-1.16666667	-3.5	0	1	0
		3.33333334	-4.0	0	0	1
2.83333333	-.66666667	-1.16666667	7.5	1.10	0	0
1.0	-.23529412	- .41176471	2.64705883	.35294118	0	0
	2.67647058	-1.44117648	-1.73529411	.23529412	1	0
	1.0	- .53846154	- .64835164	.08791208	.37362637	0
		2.07692307	-1.84615382	.53846153	.53846154	1
		1.0	- .88888889	.25925926	.25925926	.48148148
				.51322751	.22751322	.25925925
					.51322751	.25925925
						.48148148

Subtracting 1/3 from the inverse above and denoting as D', we get:

$$D' = \begin{Bmatrix} .1798418 & -.10582011 & -.07407408 \\ & .17989418 & -.07407408 \\ & & .14814815 \end{Bmatrix}$$

Estimates of the Treatment Effects are as follows:

$b_3 = -\ .88888889$

$b_2 = -\ .88888889(.53846154) + (-.64835164)\ (1.0)$

$= -\ 1.12698412$

$b_1 = -\ .88888889(.53846154) + (-.64835164)$
$(.23529412) + (2.64705883)(1.0)$

$= \ 2.01587302$

SS Treat(adj for Blocks) = ΣCPID

$= 7.5(2.64705883) + (-1.73529411)(-.64835164)$

$+ (-1.84615383)(-.88888889)$

$= 22.61904762$

AOV

Source		d.f.	S.S.	M.S.	F
Total		12	30.00	—	
Block		4	1.67	—	
Block(adj)	4		1.09	.27	< 1
Treat (adj)		2	22.62	— 11.31	12.26
Treat	2		23.20	—	
Error		6	5.71	.95	

$$\text{Treat SS} = \frac{20^2}{4} + \frac{8^2}{4} + \frac{11^2}{5} - \frac{39^2}{13} = 23.20$$

Block(adj) SS = Block SS + Treat(adj) SS − Treat SS
= 24.29 − 23.20 = 1.09

Since there is a strong evidence of treatment differences we may test the following hypothesis using the elements of D' and a t-test:

$$H_0: \rho_1 - \rho_3 = 0$$

$$t = \frac{b_1 - b_3}{S_{b1-b3}} = \frac{2.01587302 + .88888889}{.67252} = 4.319$$

where $S_{b1-b3} = \sqrt{EMS\{d'_{11} + d'_{33} - 2d'_{13}\}}$

$$= \sqrt{.95[.17989418 + .14814815 - 2(-.074074080)]}$$

$$= \sqrt{.4523809655}$$

$$= .67252$$

EXAMPLE 2 (Two-Way Classification with interaction)

Date:

		ρ			Total
	1	7 2 6 5	4 2	3 7 6 4 5	51
	2	8 1 2 4	-	5	20
a	3	2 5	7 5	9 1 9 3 9 5	48
	4	-	2 2 1	20	25
	Total	42	23	79	144

No. of Observations:

		ρ 1	ρ 2	ρ 3	
	1	4	2	5	11
$\mathcal{L}$	2	4	0	1	5
	3	2	2	6	10
	4	0	3	1	4
		10	7	13	30

To get ρ (adjusted for α) we set up:

11	0	0	0	4	2	5	51	I
	5	0	0	4	0	1	20	I
		10	0	2	2	6	48	I
			4	0	3	1	25	I
				10	0	0	42	II
					7	0	23	II
						13	79	II

Doolittle part I:

11	0	0	0	4	2	5	51
1	0	0	0	4/11	2/11	5/11	51/11
	5	0	0	4	0	1	20
	1	0	0	4/5	0	1/5	4
		10	0	2	2	6	48
		1	0	2/10	2/10	6/10	48/10
			4	0	3	1	25
			1	0	3/4	1/4	25/4

At this point we eliminate $\hat{a}$ using (4) and (5)

$$c_{11} = n_{1.} - \sum^{a_1} \frac{n^2_{i1}}{n_{i.}} = 10$$

$$- \left(\frac{4}{11} \times 4 + \frac{4}{5} \times 4 + \frac{2}{10} \times 2\right) = \frac{2720}{550}$$

$$c_{12} = - \sum_{i}^{a} \frac{n_{i1}n_{i2}}{n_{i.}} = - \left(\frac{4}{11} \times 2 + \frac{2}{10} \times 2\right) = - \frac{620}{550}$$

$$g_1 = Y_{.1.} - \sum_{i}^{a} \frac{n_{i1}Y_{i..}}{n_{i.}}$$

$$= 42 - (\frac{4}{11} \times 51 + \frac{4}{5} \times 20 + 0 \times 25) = -\frac{2360}{1100}$$

and so on and so forth.

$$C\hat{P} = 9$$

$$\begin{array}{ccc|c}
\frac{2720}{550} & -\frac{620}{550} & -\frac{2100}{550} & -\frac{2360}{1100} \\
-\frac{620}{550} & \frac{4385}{1100} & -\frac{3145}{1100} & -\frac{16.085}{1100} \\
-\frac{2100}{550} & -\frac{3145}{1100} & \frac{7345}{1100} & \frac{18445}{1100}
\end{array}$$

Now add 1/3, convert to decimal and Doolittle we get:
(for convenience rounded to 2 decimals from hereon) :

5.27	— .79	—3.48	— 2.14	1	0	0
	4.31	—2.52	—14.62	0	1	0
		7.01	16.76	0	0	1
5.27	— .79	—3.48	— 2.14	1	0	0
1	— .15	— .66	— .40	.18	0	0
	4.20	—3.04	—14.94	.15	1	0
	1	— .72	— 3.55	.03	.23	0
		2.49	4.49	.76	.72	1
		1	1.80	.30	.29	.44

$$b_3 = 1.80(1) = 1.80$$

$$b_2 = 1.8(.72) + (-3.55)(1.0) = -2.24$$

$$b_1 = 1.8(.76) + (-3.55)(.15) + (-.40)\ (1) = .44$$

Then get the inverted matrix:

.43	.25	.30
.25	.44	.29
.30	.29	.40

Any row or column of the above should add to 1. Finally subtract 1/3 from each element of the foregoing matrix we get the inverse:

.09	—.07	—.02
—.07	.11	—.04
—.02	—.04	.06

Any row or column of the above should add to zero.

$$\Sigma CPID = SS\rho(\text{adj: for } \alpha) = -2.14(-.40) + (-14.94)(-3.55) + 4.49\ (1.80)$$
$$= 62.1644$$

Analyses of Variance

ρ (adj for α)

Source	d.f.	S.S.
Total	29	400.8000
α(unadj)	3	11.9045
ρ(adj for α)	2	ΣCPID = 62.1644
$\alpha\rho$	4	by subt. = 203.9811
Within Cells	20	122.7500

α (adj for ρ)

Source	d.f.	S.S.
Total	29	400.8000
ρ(unadj)	2	40.8483
α (adj for ρ)	3	α(unadj) + ρ(adj) -ρ(unadj)=33.2206
$\alpha\rho$	4	203.9811
Within Cells	20	122.7500