ANALYSIS OF DISPERSION FOR MULTIPLY CLASSIFIED DATA WITH UNEQUAL NUMBERS IN CELLS

By C. RADHAKRISHNA RAO
Indian Statistical Institute, Calcutta

1. INTRODUCTION

The method of obtaining the components of analysis of variance table for a two-way classified data with unequal numbers in the cells is fairly well known. Illustrations can be found in many books on statistical methods (Goulden, Kendall, Rao etc.). The case of multiple classification was not fully considered even for purposes of analysis of variance. The object of this paper is to provide some general computational techniques required in the analysis of multiply classified data when there are unequal numbers in cells and when more than one character is under study. Incidentally a very simple method of computing the components of the analysis of variance for a two way table is provided.

Tests in analysis of dispersion are obtained by a generalization of the corresponding tests in analysis of variance applied to a single variable. This technique due to Fisher (1930) consists in replacing the multiple variables by a linear compound for which a variance ratio test is constructed. The compounding coefficients are chosen to maximize this ratio. If $y = m_1x^1+...+m_px^p$ is a linear compound of the p variables $x^1, ..., x^p$, then an analysis of variance of y leads among others to two components one, the sum of squares due to error w and the other, due to deviation from hypothesis b. The ratio of b to w after dividing by the degrees of freedom provides the test of the given hypothesis. In terms of x^i, the expressions b and w will be quadratic forms in $m_1, ..., m_p$.

ANALYSIS OF VARIANCE OF A LINEAR COMPOUND

due to	d.f.	sum of squares
deviation from hypothesis	k	$b = \Sigma\Sigma m_i m_j B_{ij}$
error	$n-k$	$w = \Sigma\Sigma m_i m_j W_{ij}$

The ratio
$$\frac{b}{w} = \frac{\Sigma\Sigma m_i m_j B_{ij}}{\Sigma\Sigma m_i m_j W_{ij}} \qquad \dots \quad (1.0)$$

when maximized leads to the determinantal equation

$$|B-\lambda W| = 0 \qquad \dots \quad (1.1)$$

with the largest root corresponding to the maximum value of the ratio (1.0), or writing $\mu = 1/(1+\lambda)$, the determinantal equation (1.1) can be written $|W-\mu T| = 0$, where $T = B+W$. The smallest value of μ will then correspond to the largest value of λ.

Not only the smallest root μ but also the others provide tests of departure from the null hypothesis. While among individual roots the smallest provides the best test, it is not the best test available specially when the deviation from hypothesis is not concentrated in just one linear compound of the variables. If the sample size is large all the roots are worthy of consideration in setting up a test criterion. One such is the product of the roots $\mu_1\mu_2 \ldots \mu_p$ leading to the Wilks ratio $|W| \div |T|$. The product of the last $(p-r)$ roots $\mu_{r+1} \ldots \mu_p$ provides the test for the hypothesis whether the deviation from hypothesis is concentrated in just r linear functions of the p variables.

2. Analysis of dispersion

The first step in obtaining the test criteria is the computation of the matrices B and W. The formulas for computing the elements are provided by the corresponding expressions for b and w in the univariate case. Let us consider the general problem of linear hypotheses which in the univariate case is treated by the method of least squares providing the components of the analysis of variance table. The generality of the treatment of linear hypotheses by the method of least square is not sufficiently recognized in literature. For instance analysis of variance in a stochastic model where the additive effect due to a category or a classification is considered as a random variable is considered separately. In such a case the observations in a category are equicorrelated and can be chosen to be uncorrelated by an orthogonal transformation. Then the problem comes under the general theory of least squares, where all the observations do not have the same variance. The orthogonal transformation gives rise to two sets of variables with different unknown variances, only one set of variables being relevant to the problem.

In the univariate case of n independent observations $y_1, \ldots, y_n$

$$\left.\begin{aligned} E(y_i) &= a_{i1}\tau_1 + \ldots + a_{im}\tau_m \\ V(y_i) &= \sigma^2 \end{aligned}\right\} \; i = 1, \ldots, n \qquad \ldots \quad (2.1)$$

where τ_r are unknown. The hypothesis to be tested is

$$h_{i1}\tau_1 + \ldots + h_{im}\tau_m = \xi_i, \qquad i = 1, \ldots, k. \qquad \ldots \quad (2.2)$$

In the multivariate case, instead of y_i we have a vector $(x_i^1, \ldots, x_i^p)$ with a dispersion matrix Λ independent of i and

$$E(x_i^j) = a_{i1}\tau_1^j + \ldots + a_{im}\tau_m^j, \qquad j = 1, \ldots, p; \quad i = 1, \ldots, n$$

where τ_i^j are unknown and the hypothesis to be tested is

$$h_{i1}\tau_1^j + \dots + h_{im}\tau_m^j = \xi_i^j$$

$$j = 1, \dots, p; \qquad i = 1, \dots, k$$

which is a simultaneous hypothesis of the type (2.2) for all the p variables.

To obtain the expression for w we first set up the normal equations

$$\begin{aligned} g_{11}t_1 + \dots + g_{1m}t_m &= Q_1 \\ \cdot \quad \dots \quad \cdot \quad &\quad \cdot \\ g_{m1}t_1 + \dots + g_{mm}t_m &= Q_m \end{aligned} \qquad \dots \quad (2.3)$$

obtained by equating the derivatives of

$$\Sigma(y_i - a_{i1}\tau_1 - \dots - a_{im}\tau_m)^2$$

to zero. In the equations (2.3)

$$g_{ij} = \sum_{r=1}^{n} a_{ir}a_{jr}, \quad Q_i = \sum_{r=1}^{n} a_{ir}y_r. \qquad \dots \quad (2.4)$$

The expression for the least sum of squares is

$$w = \Sigma y_r^2 - \Sigma t_r Q_r \qquad \dots \quad (2.5)$$

where t_r is a solution of (2.3). Now replacing y by $m_1x^1 + \dots + m_px^p$ in the formula (2.5) we have

$$\Sigma\Sigma m_i m_j(\Sigma x_r^i x_r^j - \Sigma t_r^i Q_r^j) = \Sigma\Sigma m_i m_j W_{ij}$$

where Q_r^j is same as Q_r with y replaced by x^j and t_r^j are the solutions of (2.3) with Q_r replaced by Q_r^j. Hence it follows that

$$W_{ij} = \Sigma x_r^i x_r^j - \Sigma t_r^i Q_r^j. \qquad \dots \quad (2.6)$$

The only computational technique needed is the solution of several sets of equations with a common matrix of equations

$$\begin{aligned} g_{11}t_1 + \dots + g_{1m}t_m &= Q_1^1, Q_1^2, \dots, Q_1^p \\ \cdot \quad \dots \quad \cdot \quad & \\ g_{m1}t_1 + \dots + g_{mm}t_m &= Q_m^1, Q_m^2, \dots, Q_m^p. \end{aligned} \qquad \dots \quad (2.7)$$

It will be shown in a later section that the actual solutions are not needed to obtain the values of $\Sigma t_r^i Q_r^j$ but in practical problems it is also necessary to obtain $t_1, \dots, t_m$ to derive the estimates of some parametric functions for a proper interpretation of the tests of significance.

The matrix proofs of the simplifying methods of computation used in the illustrations are given in an Appendix where illustrative examples are also provided. Any worker interested in carrying out the computational techniques suggested here is first advised to familiarize himself with the illustrative examples of the Appendix.

To obtain the expression for b, the sum of squares due to deviation from hypothesis there are two ways (Rao, 1946). One is to form normal equations subject to the restrictions of the hypothesis and find the least sum of squares. This gives the expression for $b+w$ from which b can be obtained by subtracting w, the unrestricted minimum. The multivariate method of obtaining $(B_{ij}+W_{ij})$ is thus same as that of W_{ij} except that equations (2.7) to be solved correspond to normal equations obtained with restrictions imposed by the hypothesis.

Another method of computing b is to find the estimates of the deviations

$$h_{i1}\tau_1+ \dots +h_{im}\tau_m-\xi_i, \qquad i=1, \dots, k$$

by substituting the solutions of the unrestricted normal equations provided, of course, the linear parametric functions in the hypothesis are estimable (Rao, 1952). If $d_1, \dots, d_k$ are the estimates with the associated variance-covariance matrix $(S_{ij})\sigma^2$ then the sum of squares due to deviation from hypothesis is provided by the expression,

$$\Sigma\Sigma C_{ij}d_id_j \qquad \dots \quad (2.8)$$

where (C_{ij}) is the matrix reciprocal to (S_{ij}). Both the methods lead to the same expression (Rao, 1946). A method of computing the expression (2.8) is to set up the equations

$$q_1S_{11}+ \dots +q_kS_{1k} = d_1$$
$$. \qquad \dots \qquad . \qquad .$$
$$q_1S_{k1}+ \dots +q_kS_{kk} = d_k$$

and obtain a solution $(q_1, \dots, q_k)$. The expression (2.8) is equal to $(q_1d_1+ \dots +q_kd_k)$. In the multivariate case the equations are

$$q_1S_{11}+ \dots +q_kS_{1k} = d_1^1, \dots, d_1^p$$
$$. \qquad \dots \qquad .$$
$$q_1S_{k1}+ \dots +q_kS_{kk} = d_k^1, \dots, d_k^p$$

and the expression for B_{ij} is

$$q_1^id_1^j+ \dots +q_k^id_k^j \qquad \dots \quad (2.9)$$

where q_r^i are the solutions to the i-th set of normal equations. As in an earlier situation (2.7) the actual solution q_i are not needed for computing the expressions (2.8) or (2.9). The method is illustrated in Sections 3, 4 while a proof is given in the Appendix.

Thus no new techniques are involved in setting up an analysis of dispersion table. Once the expressions for the components of an analysis of variance table are deduced the formulas for analysis of dispersion are automatically supplied by substitution of a linear compound and recognising the elements of the quadratic form. Given below are some multivariate problems which are generalizations of the univariate problems and in the solution of which univariate techniques are useful.

Univariate tests	*Multivariate generalization*
1. Hypothetical mean, based on a single sample	Simultaneously for p-variables, based on a single p-variate sample
2. Equality of means, two samples	Simultaneously for p-variables, two p-variate samples
3. Multi-classification data, main effects and interactions	Main effects and interactions simultaneously for p-variables
4. Intraclass Correlation	Familial Correlations
5. Hypothesis about regression functions etc.	Canonical regressions and simultaneous tests for p-variables
6. Simple, partial and multiple correlations	Canonical correlations
7. Equality of means of p equi-correlated variables	Equality of means of p-variables not necessarily equi-correlated

3. Two-way classification

Let us suppose that there are rs cells defined by two factors A and B at levels and s respectively. The number of observations in (k, m)th cell is denoted by n_{km} and the total of the observations for the i-th character by x^i_{km}. The marginal totals are obtained by the usual summation notation, $x_k^i.$, $x^i_{.m}$, $n_{k.}$, $n_{.m}$, $x^i_{..}$, and so on. If the numbers in the cells are all equal to n, then the analysis of dispersion is as shown in Table 1.

TABLE 1. ANALYSIS OF DISPERSION FOR EQUAL NUMBERS IN CELLS

due to	d.f.	S_{ij}
A	$r-1$	$1\ \Sigma x^i_{k.} x^j_{k.} - \text{c.f.}$
B	$s-1$	$\Sigma x^i_{.m} x^j_{.m} - \text{c.f.}$
AB	$(r-1)(s-1)$	—
between cells	$rs-1$	$\frac{1}{n} \Sigma\Sigma x^i_{km} x^j_{km} - \text{c.f.}$
within cells	—	—
total	$rsn-1$	$\Sigma\Sigma\Sigma x^i_{kmt} x^j_{kmt}$

c.f. $= x^i_{..} x^j_{..}/nrs$ — obtained by subtraction

The case of unequal numbers can be best illustrated by an example. The method followed appears to be the simplest computational technique of analysis of variance or dispersion for a two-way classified data. For illustrative purposes the example on nasal length (nl) given in Rao (1952), p. 95, is taken with further observations on nasal breadth (nb). The data relate to measurements on nl and nb taken by 3 different observers O_1, O_2, O_3 on different sets of skulls recovered by them from three different strata S_1, S_2, S_3.

Table 2 gives, for $x^1 = nl$, $x^2 = nb$ the cell totals x_{km}, means $\bar{x}_{km}$, numbers n_{km}, marginal totals $x_{k.}$, $x_{.m}$ etc. The formulas and the computations for the Sums of Squares and Products (SP matrix) for 'between cells', 'strata ignoring observers' (obtained from strata totals only) and 'observers ignoring strata' are also provided near the table to make the formulas more explanatory. In practice the entries for all SP matrices need only be recorded in the final table of analysis of dispersion. All the formulas are familiar except for the complication introduced by two variables. The extension to several variables is simple. The within cell SP matrix is obtained by subtracting from the total, the between cell matrix.

TABLE 2. TOTALS AND MEANS FOR CELLS AND MARGINALS

strata		S_1		S_2		S_3		all strata	
		total	mean	total	mean	total	mean	total	mean
observers		11		12		13		1.	
	x^1	1071.00	51.00000	1572.48	49.14000	913.50	50.75000	3550.98	50.00831
O_1	x^2	472.19	22.40000	728.62	22.76937	430.76	23.93111	1631.67	22.98127
	n	21		32		18		71	
		21		22		23		2.	
	x^1	1966.86	46.83000	2315.40	45.40000	1721.52	47.82000	6003.78	46.54093
O_2	x^2	890.64	21.20571	1048.05	20.55000	777.72	21.60333	2716.41	21.05744
	n	42		51		36		129	
		31		32		33		3.	
	x^1	1219.00	48.76000	2091.60	46.48000	1849.20	46.23000	5159.80	46.90727
O_3	x^2	568.00	22.72000	944.79	20.99533	844.55	21.11375	2358.33	21.43936
	n	25		45		40		110	
		.1		.2		.3		..	
	x^1	4256.86	48.37341	5979.48	46.71469	4484.22	47.70447	14720.56	47.485677
all observers	x^2	1931.02	21.94364	2721.46	21.26141	2053.03	21.84075	6706.41	21.633581
	n	88		128		94		310	

The suffix (k,m) is indicated in the corner of a cell. The cell totals for nl were reconstructed from published mean values correct to two decimal places. $x^1 = nl$, $x^2 = nb$.

Computations and notes based on Table 2

(i) *Correction factors*

$$C^{11} = x^1_{..}\bar{x}^1_{..} = 14720.50 \times 47.485077 = 699015.76$$
$$C^{12} = x^1_{..}\bar{x}^2_{..} = \bar{x}^1_{..}x^2_{..} = 318458.43$$
$$C^{22} = x^2_{..}\bar{x}^2_{..} = 145083.66$$

(ii) *SP matrix for strata ignoring observers*
(only marginal totals are involved)

$$S_{11} = \Sigma x^1_{.l}\bar{x}^1_{.l} - C^{11}$$
$$= (1256.86)(48.37341)+\ldots+(4484.22)(47.70447)-C^{11}$$
$$= 699165.73-699015.76 = 149.97$$
$$S_{12} = \Sigma x^1_{.l}\bar{x}^2_{.l} - C^{12} = \Sigma\bar{x}^1_{.l}x^2_{.l} - C^{12} = 66.05$$
$$S_{22} = \Sigma x^2_{.l}\bar{x}^2_{.l} - C^{22} = 30.80$$

(iii) *SP matrix for observers ignoring strata*
(only marginal totals are involved)

$$S_{11} = \Sigma x^1_{l.}\bar{x}^1_{l.} - C^{11} = (3556.98)(50.09831)+\ldots-C^{11} = 636.56$$
$$S_{12} = \Sigma x^1_{l.}\bar{x}^2_{l.} - C^{12} = \Sigma\bar{x}^1_{l.}x^2_{l.} - C^{12} = 332.53$$
$$S_{22} = \Sigma x^2_{l.}\bar{x}^2_{l.} - C^{22} = 175.91$$

(iv) *SP matrix between cells*

$$S_{11} = \Sigma\Sigma x^1_{km}\bar{x}^1_{km} - C^{11}$$
$$= 1071.00(51.00000)+1966.86(46.830)+\ldots+1849.20(46.230)-C^{11}$$
$$= 699947.62-699015.76 = 931.86$$
$$S_{12} = \Sigma\Sigma x^1_{km}\bar{x}^2_{km} - C^{12} = 475.85$$
$$S_{22} = \Sigma\Sigma x^2_{km}\bar{x}^2_{km} - C^{22} = 280.12$$

(v) All the formulas make use of the mean values, which are needed in any case for the final interpretation. If sufficient significant figures are not recorded, the mean may be replaced by the total and then the square or the product is divided by the number of observations making up the total.

(vi) For the univariate case consider only the first formula in each of the categories (i), (ii), (iii), (iv). For more than two variables the formulas for S_{ij} are similar.

3a. *Computation of Interaction and Main effects:*

Assuming that effects of observers and strata are additive we can write down normal equations for observer effects O_1, O_2, O_3 and strata effects S_1, S_2, S_3. The value of S_3 can be arbitrarily put as zero so that there are only 5 equations as in Table 3. In general the value of one level of one of the factors can be taken to be zero so that the number of equations is $(r+s-1)$, one less than the sum of the levels of the two factors. The method of writing these equations is very simple. In the block for characters we have the marginal totals. The marginal totals 3556.98 and 1631.67 for *nl* and *nb* are based on 71 observations of O_1, none (zero) of O_2, O_3 and 21 of S_1 and 32 of S_2. This is the equation for O_1, similarly equations for O_2, O_3, S_1 and S_2 are written. This is sufficient if only the sum of squares and products of interaction and main effects are required. If the aim is also to obtain estimates of all differences in the levels of main

TABLE 3. NORMAL EQUATIONS, CONDENSED GAUSS-DOOLITTLE SCHEME WITH AN ADJOINED UNIT MATRIX

equation	observers			strata		marginal totals for characters		unit matrix for estimation of constants etc.				
	O_1	O_2	O_3	S_1	S_2	nl	nb	O_1'	O_2'	O_3'	S_1'	S_2'
O_1	71	.	.	21	32	3556.98	1631.67	1	.	.	.	.
O_2	.	129	.	42	51	6003.78	2716.41	.	1	.	.	.
O_3	.	.	110	25	43	5159.80	2358.33	.	.	1	.	.
S_1	21	42	25	88	.	4256.86	1031.92	.	.	.	1	.
S_2	32	51	43	.	128	5979.48	2721.46	.	.	.	.	1
$O_{1\cdot 0}$	71	.	.	21	32	3556.98	1631.67	1	.	.	.	.
$O_{1\cdot 1}$	1	.	.	0.295775	0.450704	50.098310	22.981268	.0140845	.	.	.	.
$O_{2\cdot 0}$		129	.	42	51	6003.78	2716.41	.	1	.	.	.
$O_{2\cdot 1}$		1	.	0.325581	0.395349	46.540930	21.057442	.	.0077519	.	.	.
$O_{3\cdot 0}$			110	25	43	5159.80	2358.33	.	.	1	.	.
$O_{3\cdot 1}$			1	0.227273	0.409091	46.907273	21.439364	.	.	.0090909	.	.
$S_{1\cdot 0}$				62.432498	−36.296717	77.394605	28.916708	−0.295775	−0.325581	−0.227273	1	
$S_{1\cdot 1}$				1	−.581375	1.239652	0.463168	− .0047375	− .0052149	− .0036403	.0160173	.
$S_{2\cdot 0}$					53.903574	−63.085337	−35.830020	−0.622000	−0.584631	−0.541221	0.581375	1
$S_{2\cdot 1}$					1	− 1.170337	− 0.664706	− .011551	−0.010846	−0.010041	0.010785	.0185516

The sum check column has been omitted to accommodate the table in a single printed page.

(i) The normal equations for O_1, O_2, O_3, are already in a swept out form. Only division by the pivotal element is needed. This is done in the second row block. Many entries are repeated here which could be omitted. To avoid confusion they may be retained. The two equations S_1 and S_2 are reduced in the third row block. Because of the simplicity in the reduction of the first set of equations it is *always* better to write the equations for the factor with the larger number of levels first.

(ii) The method of reduction employed above is the Gauss-Doolittle form. The equivalent form of square root can also be used. See the appendix.

(iii) For obtaining the components of analysis of variance (in the case of a single variable) or analysis of dispersion (for more variables as in this example) we need only the SP matrix for 'strata eliminating observers' as shown in Table 4. The rest of the entries in Table 4 are either taken from Table 2 or obtained by subtraction. To obtain the SP matrix due to 'strata eliminating observers' the values involved are found in the columns for nl and nb and the rows in the third row block containing reduced S-equations. Each element of the SP matrix is obtained as the sum of products of elements in rows (i.o) and (i.1), the summation being over all i in row block for reduced S. Thus

$$S_{11} = 77.394605(1.239652)+(-63.085337)(-1.170337) = 169.7734$$
$$S_{12} = 77.394605(0.463168)+(-63.085337)(-0.664706) = 77.7799$$
$$S_{22} = 28.916708(0.463168)+(-35.830020)(-0.664706) = 37.2097$$

The generalization to more variables is obvious. The reduction of a unit matrix is unnecessary if the object is just to obtain the components of the analysis of dispersion table.

factors, when interaction is not significant and also their variances and covariances, it is convenient to adjoin a unit matrix and proceed with the method of pivotal condensation by using Gauss-Doolittle technique or square root.

The computations of Table 3 and the expressions derived from them are quite explanatory. With the help of the results of Table 2 and *only* the SP matrix for 'strata eliminating the observers' calculated in Table 3, the analysis of dispersion Table 4 is obtained. The general method followed here can be recommended whatever may be the levels of A and B provided in Table 3, we write the equations for the factor with higher levels first.

TABLE 4. ANALYSIS OF DISPERSION

due to	d.f.	x^1x^1	x^1x^2	x^2x^2	x^1x^1	x^1x^2	x^2x^2	d.f.	due to
observers ignoring strata from Table 2	2	636.56	332.53	175.91	636.36–	344.20–	182.32–	2–	observers eliminating strata
†strata eliminating observers from Table 3	2	169.77	77.78	37.21	169.97	66.05	30.80	2	strata ignoring observers from Table 2
interaction $O \times S$	4–	125.53–	65.54–	67.00–*	125.53	65.54	67.00	4	interaction $O \times S$
between cells from Table 2	8	931.86	475.85	280.12	931.86	475.85	280.12	8	between cells from Table 2
within cells	301–	4466.90–	1792.16–	3115.26–					
total	309	5398.76	2268.01	3395.38					

–All these quantities are obtained by subtraction from appropriate totals.

* The interaction $O \times S$ is first obtained by subtraction and then carried over to the right block to compute the SP matrix for 'observers eliminating strata' by subtraction.

† Besides the familiar computations based on the marginal totals of Table 2, the extra computations needed are for 'strata eliminating observers' only. The rest are obtained by subtraction.

3b. *Multivariate tests:*

Interaction

$$\Lambda = \frac{|W|}{|W+B|} = \frac{\begin{vmatrix} 4466.90 & 1792.16 \\ 1792.16 & 3115.26 \end{vmatrix}}{\begin{vmatrix} 4586.43 & 1857.70 \\ 1857.70 & 3182.20 \end{vmatrix}} = \frac{10703717}{11144103}$$

$$= 0.960477$$

$$\chi^2 = -\left(n - \frac{p+q+1}{2}\right) \log_e \Lambda \text{ with } pq \text{ degrees of freedom*}$$

$$= -\left(305 - \frac{2+4+1}{2}\right) \log_e(0.960477) = 12.16$$

The value of $\chi^2 = 12.16$ is not significant for 8 degrees of freedom. We can then proceed to test for stratum differences by choosing the matrix for 'strata elimiating observers'.

* In the formula for χ^2, p stands for the number of variables, q the degrees of freedom for the component to be tested and n, the total degrees of freedom for error and the component to be tested.

Strata and Observers

$$\Lambda = \frac{|W'|}{|W'+B|} = \frac{10703717}{\begin{vmatrix} 4636.67 & 1869.94 \\ 1869.94 & 3152.47 \end{vmatrix}} = \frac{10703717}{11120287}$$

$$\chi^2 = -\left(303 - \frac{2+2+1}{2}\right)\log_e(.962539) = 11.47.$$

This is significant for 4 degrees of freedom. Similarly for observers $\chi^2 = 42.50$ which is significant for 4 degrees of freedom.

Strata using interaction as error

In problems where interaction is the appropriate error

$$\Lambda = \frac{\begin{vmatrix} 125.53 & 65.54 \\ 65.54 & 67.00 \end{vmatrix}}{\begin{vmatrix} 295.30 & 143.32 \\ 143.32 & 104.21 \end{vmatrix}} = \frac{4115.02}{10232.59}$$

$$\chi^2 = -\left(6 - \frac{2+2+1}{2}\right)\log_e\Lambda = 3.19$$

which is small for 4 degrees of freedom. Since the degrees of freedom are small the Beta approximation given by the author (Rao, 1952) has to be used. But the χ^2 is so small that even the true probability is expected to be far above 5%.

3c. *Estimation of constants and their variances-covariances:* It was seen that for analysis of variance or dispersion it is not necessary to actually solve the normal equations of Table 3 or use the elements of reduced unit matrix. If, however, we are interested in judging the individual differences between the levels of each factor it is necessary to obtain a solution and also derive the variance of the estimated differences. The equations can be obtained by back solution and also the inverse matrix by a series of back solutions. But with a unit matrix adjoined to the normal equations and reduced simultaneously there are some advantages as discussed in the Appendix. The constants can be computed without a back solution and the matrix of the variances-covariances can be obtained very simply.

We define the product $(x).(y)$ of two columns (x) and (y) in the Gauss-Doolittle scheme as follows. It is the sum of products of elements one from row (i .0) in column x (or y) and another from row (i.1) in column y (or x), summation being over all i. With this definition the estimates of the constants and their variances-covariances can be expressed as column products. The estimates of the constants and their variances-covariances are found from Table 3 as follows.

$$O_1 = (nl).(O'_1) = 3556.98(.014085)+6003.78(0)+\ldots+(-63.0853)(-.011551)$$
$$=50.4586$$

$$\ldots$$

$$S_1 = (nl).(S'_1) = 3556.98(0)+\ldots+(-63.0853)(.018552) = -1.170337$$

$Cov\,(O_1, O_2) = (O'_1).(O'_2)\sigma^2_{nl}$

$(O'_1).(O'_2) = 1(0)+0(.007752)+\ldots+(-0.622660)(-0.010846) = .008296$

Similarly for nb. The constants can be computed in any order because of the reduced form of the unit matrix.

THE CONSTANTS AND THEIR VARIANCES-COVARIANCES

	constants		variance-covariance matrix without multiplying by the variance		
	nl	nb			
O_1	50.4586	23.2573	.022677	.008296	.007329
O_2	46.8210	21.2954	.008296	.015791	.007055
O_3	47.2595	21.6941	.007329	.007055	.022799
S_1	.5593	0.0767	.022287	.010786	0
S_2	−1.1703	−0.6647	.010786	.018552	0
S_3	0	0	0	0	0

These constants by themselves have no meaning but are useful in estimating the comparisons. For instance to test whether the difference between O_1 and O_3 is significant for nl we find the difference $O_1-O_3 = 50.4586-47.2595 = 3.1991$ with variance

$$V(O_1)-2\,Cov\,(O_1, O_3)+V(O_3) = (.022677-2(.007329)+.022799)\sigma^2_{nl} = .030818\sigma^2_{nl}.$$

The values of the variances (V) and covariances (Cov) are as in the above table.

The t test can be made by substituting the estimate of σ^2_{nl} from Table 4, in the ratio $3.1991/\sqrt{.030818\sigma^2_{nl}}$. Similarly for nb

$$S_1-S_3 = 0.0767-0 = 0.0767$$

$$V(S_1-S_3) = (.022287-2(0)-0)\sigma^2_{nb} = .022287\sigma^2_{nb}.$$

Thus tests of all differences can be carried out.

4. Three way classification

The analysis follows usual lines if cell numbers are equal. The SP matrices for all the main effects and interactions add up to the total. Otherwise computations become tedious. When only a few observations are missing, the cell averages may be substituted for the missing values and the analysis can be carried out as in the case of equal numbers without serious error. But when the cell numbers differ very much from one another, it is not known how good the above approximate treatment is. If an accurate analysis of the material is desired the following technique may be useful.

As in two way classification, the first step is to obtain analysis of total dispersion as between and within cells. The next step is to obtain the SP matrix due to the triple factor interaction.

4a. *Triple factor interaction:* Let us recall that a component of the triple factor interaction can be symbolically represented as

$$(A_i-A_{i+1})(B_j-B_{j+1})(C_k-C_{k+1}) \qquad \dots \quad (4.1)$$

which is a linear compound of eight combinations of three factors A, B, C. By letting i, j, k run through the different levels p, q, r of A, B, C, a total of $(p-1)(q-1)(r-1)$ independent components can be generated. A component like (4.1) for any observed variable can be estimated by inserting the mean values of the cells defined by the combinations. Since the estimates are linear combinations of the cell averages, the variance-covariance matrix of these estimates can be found and the second method of calculating the SP matrix explained in Section 2 may be used. The method is illustrated with the following example.

TABLE 5. THREE WAY CLASSIFICATION. MEANS FOR x^1, x^2 IN THE CELLS, NUMBERS AND RECIPROCALS IN BRACKETS

method	machine setting	type of flour							
		1		2		3		4	
		x^1	x^2	x^1	x^2	x^1	x^2	x^1	x^2
M_1	S_1	9.4 (4)	11.5 (.25)	2.6 (4)	3.1 (.25)	12.3 (5)	10.8 (.2)	4.6 (5)	0.2 (.2)
	S_2	9.6 (2)	6.4 (.5)	3.1 (4)	−6.1 (.25)	13.0 (4)	6.1 (.25)	4.3 (2)	7.2 (.5)
	S_3	9.6 (1)	5.3 (1)	2.7 (2)	−3.0 (.5)	13.4 (4)	14.3 (.25)	1.8 (5)	−5.1 (.2)
M_2	S_1	13.7 (1)	18.7 (1)	21.6 (4)	24.5 (.25)	19.4 (4)	27.7 (.25)	13.5 (5)	17.9 (.2)
	S_2	12.7 (4)	12.0 (.25)	22.6 (4)	22.1 (.25)	20.6 (5)	24.2 (.2)	10.4 (4)	11.1 (.25)
	S_3	12.6 (2)	8.6 (.5)	21.8 (5)	16.8 (.25)	20.9 (4)	12.3 (.25)	6.8 (4)	3.3 (.25)

The six components of the interaction $F \times S \times M$ can be generated from the scheme

$$\begin{matrix} F_1-F_2 \\ F_2-F_3 \\ F_3-F_4 \end{matrix} \times \begin{matrix} S_1-S_2 \\ S_2-S_3 \end{matrix} \times M_1-M_2$$

For instance the first one $(F_1-F_2)\,(S_1-S_2)\,(M_1-M_2)$ or

$$F_1S_1M_1+F_2S_2M_1-F_1S_2M_1-F_2S_1M_1-F_1S_1M_2-F_2S_2M_2+F_1S_2M_2+F_2S_1M_2$$

has the estimate

$$\{(9.4+3.1)-(9.6+2.6)\}-\{(13.7+22.6)-(12.7+21.6)\}$$

To find the covariance (apart from a multiplier σ^2) of two components the reciprocals of numbers on which the averages are based are combined with coefficients which are the product of the coefficients of the averages in the two components. For instance the variance of the above estimate is

$$(.25+.25+.5+.25+1+.25+.25+.25)\sigma^2 = 3\sigma^2.$$

Similarly the covariance of the estimates of

$$(F_1-F_2)(S_1-S_2)(M_1-M_2) \text{ and } (F_2-F_3)(S_1-S_2)(M_1-M_2)$$

is (we need only consider the common terms and the reciprocals of the averages corresponding to them)

$$(-.25-.25-.25-.25)\sigma^2 = -1\times\sigma^2$$

and so on. The entire matrix of variances -covariances and the estimates and also its square root reduction is shown in Table 6.

TABLE 6. SQUARE ROOT METHOD FOR THE EVALUATION OF $F\times S\times M$

row	variance-covariance matrix						x^1	x^2
1	3.00	−1.25	−1.00	.50	.	.	− 1.7	− 8.4
2		3.50	.50	1.25	.	.	0.3	6.1
3			1.00	− .95	− .90	.45		5.6
4				2.20	.45	− .05	− 1.3	11.7
5					2.05	−1.20	3.3	15.0
6						2.15	2.0	−24.6
	square root reduction							
1.1	1.73205	− .72169	− .57735	.28868	.	.	− .08150	− 4.84974
2.1		1.72602	.04828	− .60349	.	.	− .23058	1.50635
3.1			1.25074	− .00300	− .71957	.35079	− .44394	2.18052
4.1				1.17840	.01366	− .62202	−1.21098	13.00303
5.1					1.23775	− .75347	2.42141	13.24201
6.1						1.03243	3.12044	− 7.08837

The six components chosen are in the order—

$1=(F_1-F_2)(S_1-S_2)(M_1-M_2)$, $2=(F_1-F_2)(S_2-S_3)(M_1-M_2)$, $3=(F_2-F_3)(S_1-S_2)(M_1-M_2)$,
$4=(F_2-F_3)(S_2-S_3)(M_1-M_2)$, $5=(F_3-F_1)(S_1-S_2)(M_1-M_2)$, $6=(F_3-F_4)(S_2-S_3)(M_1-M_2)$.

The SP matrix due to $F\times S\times M$ can be obtained by column multiplication of elements in the rows (1.1) to (6.1). Thus

$$S_{11}=(x^1).(x^1)=(-.08150)^2+(-.23058)^2+\ldots+(3.12044)^2=18.330$$

$$S_{12}=(x^1).(x^2)=(-.08150)(-4.84974)+\ldots+(3.12044)(-7.08837)=-2.427$$

$$S_{22}=(x^2).(x^2)=-(4.84074)^2+\ldots+(-7.08837)^2=425.2.$$

In the reduction of the matrix in Table 6 it is better to keep a sum check column. This is omitted to accommodate the table within the printed page.

This is a general method of computing the triple factor interaction whatever may be the number of levels of the factors. But if one of the factors has two levels as in the above example a simpler method is available. In fact by using this method, not only the triple factor interaction but two first order (two factor) interactions are simultaneously obtained. In practice the triple factor interaction may be split into different sets of degrees of freedom (in a meaningful way, not necessarily orthogonal) such that in each set only one comparison of at least one of the factors is involved. For instance if there had been three methods in the above problem (M_1, M_2, M_3), we could calculate the triple factor interaction in two sets each with six degrees of freedom comprising the components

$$\begin{matrix} F_1-F_2 \\ F_2-F_3 \\ F_3-F_4 \end{matrix} \times \begin{matrix} S_1-S_2 \\ S_2-S_3 \end{matrix} \times M_1-M_2$$

in the first set $(FSM)_1$, and

$$\begin{matrix} F_1-F_2 \\ F_2-F_3 \\ F_3-F_4 \end{matrix} \times \begin{matrix} S_1-S_2 \\ S_2-S_3 \end{matrix} \times M_2-M_3$$

in the second set $(FSM)_2$. It could be split in any meaningful way. What is needed is that only one comparison of one of the factors is taken into account. In the above illustration we first construct a two way table $F \times S$ containing the differences of the averages between M_1 and M_2 and their weights (the harmonic mean of the numbers for M_1 and M_2) from Table 5. For instance in the cell F_1S_1, the difference between M_2 and M_1 is $13.7-9.4 = 4.3$ with the weight $(4 \times 1)/(4+1) = 0.8$. We have the following table.

TABLE 7. DIFFERENCES (M_2-M_1) WITH WEIGHTS

	F_1		F_2		F_3		F_4	
	x^1	x^2	x^1	x^2	x^1	x^2	x^1	x^2
S_1	4.3	7.2	10.0	21.4	7.1	16.0	8.0	17.7
wt	0.8		2.0		2.2		2.5	
S_2	3.1	5.6	10.5	28.2	7.6	18.1	6.1	3.9
wt	1.3		2.0		2.2		1.3	
S_3	3.0	3.3	10.1	10.8	6.5	-2.0	5.0	8.4
wt	0.7		1.4		2.0		2.2	

Table 7 is treated exactly in the same way as the two way data of Table 2 with mean values and weights as numbers. The computation Table 3 and analysis of dispersion Table 4, for main effects and interaction may be followed. The 'main effect S eliminating F' is now the interaction SM and the 'main effect F eliminating S' is the interaction FM and the 'interaction FS' is the interaction FSM. The between cell sum of products' is obtained from the entries in the cells and the weights

$$B_{rs} = \Sigma\Sigma w_{ij} d^r_{ij} d^s_{ij} - \frac{(\Sigma\Sigma w_{ij} d^r_{ij})(\Sigma\Sigma w_{ij} d^s_{ij})}{\Sigma\Sigma w_{ij}}.$$

We will just write down the normal equations needed without carrying out the computations. The interested reader may follow the Gauss-Doolittle or square root methods and verify that FSM has the same value as that obtained in Table 6 correct up to the significant figures retained.

TABLE 8. NORMAL EQUATIONS BASED ON TABLE 7 TO BE REDUCED AS IN TABLE 3 FOR THE COMPUTATION OF $F\times M$, $S\times M$ AND $F\times S\times M$

F_1	F_2	F_3	F_4	S_1	S_2	S_3	x^1	x^2
2.8	.	.	.	0.8	1.3	0.7	9.57	23.35
.	5.4	.	.	2.0	2.0	1.4	103.74	126.92
.	.	6.4	.	2.2	2.2	2.0	49.34	73.00
.	.	.	6.0	2.5	1.3	2.2	41.18	67.60
0.8	2.0	2.2	2.5	7.5	.	.	70.31	129.90
1.3	2.0	2.2	1.3	.	6.8	.	67.68	108.67
0.7	1.4	2.0	2.2	.	.	6.3	56.84	44.51

The column and row (last) for S_3 may be omitted because in pivotal condensation the last row will be zero and this has no effect on the SP matrix. But as mentioned in the Appendix this may be retained for purposes of an additional check.

4b. *Two factor interactions:* The two factor interactions are meaningful only when the triple factor interaction is non-existent or not found to be significant on the basis of a test. It is seen that in the situation where only one comparison of a factor is considered two of the two factor interactions have been computed along with the triple factor interaction. To calculate FS we may break it up into sets of components $F(S_1-S_3)$ and $F(S_1-2S_2+S_3)$ and obtain the SP matrices for them individually so that the interaction FS will appear as $F(S_1-S_3)$ with 3 d.f. and $F(S_1-2S_2+S_3)$ with the other 3 d.f. To obtain the interaction $F(S_1-S_3)$ or $F(S_1-2S_2+S_3)$ it is necessary to form a two way table $F\times M$ containing the values of S_1-S_3 (or $S_1-2S_2+S_3$) and the weights (reciprocals of variances) in the cells. The 'main effect of F eliminating M' calculated from such a table by the method of Table 3 is the interaction $F(S_1-S_3)$ or $F(S_1-2S_2+S_3)$.

If, however, the SP matrix for $F\times S$ as a whole has to be obtained the more elaborate method outlined below may have to be followed.

The variance-covariance matrix and the components of interaction $F\times S$ are computed for each level of M (M can be at any number of levels). For the levels M_1 and M_2 the matrices are as follows.

TABLE 9. MATRICES FOR THE EVALUATION OF THE TWO FACTOR INTERACTION $F\times S$

	variance-covariance matrix						components of $F\times S$ z^1	z^2	
	1.25	−.75	−.50	.25	.	.	0.3	−4.1	$(F_1-F_2)(S_1-S_3)$
		2.25	.25	−.75	.	.	−.4	4.2	$(F_1-F_2)(S_2-S_3)$
M_1 :			.95	−.50	−.45	.25	.2	4.5	$(F_1-F_3)(S_1-S_3)$
				1.25	.25	−.50	−.2	5.1	$(F_1-F_3)(S_2-S_3)$
					1.15	−.75	−1.0	11.7	$(F_1-F_4)(S_1-S_3)$
						1.20	−1.9	−20.5	$(F_1-F_4)(S_2-S_3)$
	1.75	−.50	−.50	.25	.	.	2.0	4.3	
		1.25	.25	−.50	.	.	−.7	−1.9	
M_2 :			.95	−.45	−.45	.20	.2	−1.1	
				.95	.20	−.45	1.1	−6.6	
					.90	−.45	−4.3	−3.3	
						.95	−3.9	4.1	

It may be observed that Table 6 used for the computation of the triple factor interaction can be deduced by adding the variance-covariance matrices and taking differences (M_2-M_1) of the components of $F\times S$. In computing $F\times S$, the effect of $F\times S\times M$ is not considered so that the components of $F\times S$ are considered the same for each level of M. So we now need a method of combining them first taking into account the fact that the variance-covariance matrix is different for different levels of M. A general method in such a situation is provided by the least square technique in the correlated case. The inverse of the variance-covariance matrix is found for each level of M. Let $(C_{ij})_1$ be the inverse for M_1. Pre-multiplying the components for z^1, z^2 by $(C_{ij})_1$ we have two new columns represented by $(L_{rs})_1$. Similarly we have $(C_{ij})_2$ and $(L_{rs})_2$. The normal equations for the six components are obtained by addition $(C_{ij}) = (C_{ij})_1+(C_{ij})_2$ for the left hand side matrix and $(L_{rs}) = (L_{rs})_1+(L_{rs})_2$ for the right hand side. The computational technique of Table 6 is used with the matrices (C_{ij}) and (L_{rs}) to obtain the SP matrix for the interaction $F\times S$. Instead of the square root the Gauss-Doolittle method can also be used.

4c. *Main effects:* The main effects can be properly interpreted only when all the interactions do not exist. This can be judged by testing the significance of the interactions calculated earlier. For purposes of obtaining the main effects we write the expectation of the cell defined by $F_iS_jM_k$ as $F_i+S_j+M_k$ (the sum of three effects) and obtain normal equations for the constants (F_1, F_2, F_3, F_4; S_1, S_2, S_3; M_1, M_2).

Experience has shown that for the computational technique followed here it is convenient to write the equations for the factor with the largest number of levels first (here F) and for the factor with the next largest number of levels last (here S). In the following illustration we, however, follow a different order. The normal equations are written using the marginal totals for each level of a factor. The distribution of the total number for a level of a factor, among the levels of the various factors constitutes the left hand side matrix and the marginal totals for x^1 and x^2 the right hand side matrix.

TABLE 10. NORMAL EQUATIONS FOR MAIN EFFECTS
(omitting the values below the diagonal)

equation for	F_1	F_2	F_3	F_4	S_1	S_2	S_3	M_1	M_2	marginal total x^1	marginal total x^2	check x^1+x^2
F_1	14	.	.	.	5	6	3	7	7	156.1	148.0	304.1
F_2		23	.	.	8	8	7	10	13	314.0	252.4	566.4
F_3			26	.	9	9	8	13	13	427.3	416.6	843.9
F_4				25	10	6	9	12	13	176.9	137.0	313.9
S_1					32	.	.	18	14	377.7	430.4	808.1
S_2						29	.	12	17	378.0	329.0	707.0
S_3							27	12	15	318.6	194.6	513.2
M_1								42	.	298.3	171.6	469.9
M_2									46	776.0	782.4	1558.4

There is a complete check in equations of this sort. The totals of the rows and columns in each block is the same for the left hand and right hand (for x^1, x^2, x^1+x^2) matrices. The columns and rows corresponding to S_3 and M_2 may be omitted but, as we shall see, carrying them provides additional checks and we need only provide a check for the right hand side matrix. Adjoining unit matrix to Table 10, the square root method of reduction is employed. Omitting the elements of the left hand matrix the reduced matrix is as shown in Table 11.

The computations of Table 11 enable us to calculate the SP matrices due to all the main effects and also the effects of the individual levels (for comparative purposes) for each factor and their variances-covariances. We start with the last factor M in reverse order of factors F, S and M considered in Tables 10 and 11.

	Estimates x^1	x^2	*Variance-covariance matrix (without σ^2)*	
	$M_1=(M'_1).(x^1) = -0.9362$	$(M'_1).(x^2) = 13.7052$	$(M'_1)^2 = .04648$	$(M'_1).(M'_2) = 0$
	$M_2=(M'_2).(x^2) = 0$	$(M'_2).(x^2) = 0$	0	$(M'_2)^2 = 0$

From values of these constants the difference and its variance are estimated

$M_1 - M_2 \;=\; -0.9302 \qquad -13.7052 \qquad .040484$

TABLE 11. THE SQUARE ROOT REDUCTION OF TABLE 10 WITH A UNIT MATRIX

	x^1	x^2	check x^1+x^2	F'_1	F'_2	F'_3	F'_4	S'_1	S'_2	S'_3	M'_1	M'_2	sum check
F_1	41.7194	39.5546	81.2741	.26726									.26726
F_2	65.4735	52.6291	118.1026		.20851								.20851
F_3	83.8004	81.7020	165.5024			.19612							.19612
F_4	35.3800	27.4000	62.7800				.20000						.20000
S_1	− 1.3176	20.1324	18.8148	−.07024	−.07717	−.07680	−.08674	.22186					−.10009
S_2	2.2844	12.7748	15.0592	−.16522	−.14214	−.14145	−.12027	.13872	.26993				−.16043
S_3	0	0	0	0	0	0	0*	0	0				0
M_1	−46.0865	−63.9853	−110.0718	−.10190	−.08750	−.10150	−.09515	−.02517	.00722		.21560		−.12890
M_2	0	0	0	0	0	0	0	0	0		0		

*In actual computation the row will be zero to the number of significant figures retained. This is an indication that a zero row has been encountered and it should be verified theoretically at that stage.

The SP matrix for M is (when only one difference is possible)

$$S_{11} = \underset{\text{for } x^1}{(M_1 - M_2)^2} \div 0.040484 = (-9.0362)^2 \div 0.040484 = 2123.01$$

$$S_{22} = \underset{\text{for } x^2}{(M_1 - M_2)^2} \div 0.040484 = (-13.7052)^2 \div 0.040484 = 4094.04$$

$$S_{12} = \underset{\text{for } x^1}{(M_1 - M_2)}\underset{\text{for } x^2}{(M_1 - M_2)} \div 0.040484 = 2948.80.$$

For the factor which comes at the end the SP matrix can be obtained in a simpler way by using the columns (x^1), (x^2) and restricting the multiplication to the rows corresponding to M_1, M_2 (the levels can be more than two).

$$S_{11} = (x^1).(x^1) = (-46.0865)^2 + (0)^2 = 2123.96$$
$$S_{22} = (x^2).(x^2) = (-63.9853)^2 + (0)^2 = 4094.11$$
$$S_{12} = (x^1).(x^2) = 2948.86$$

which agrees with the previous calculation to four significant figures which is the accuracy expected. Because of the simplicity of the evaluation of the SP matrix without using the actual estimates it may be convenient to order the factors in such a way that the factor with the largest number of levels comes last. But the other approach of keeping such a factor first introduces simplicity in the reduction of the matrix (by square root or Gauss-Doolittle). The factor with the second largest number of levels may be kept to the last. For simplifying the computations S should have been the last factor.

Now for the factor S, we find estimates and variance-covariances by column multiplications from Table 11.

	Estimates		*Variance-covariance matrix*		
	x^1	x^2			
S_1	1.1846	7.8492	.069099	.037263	0
S_2	0.2839	2.9803	.037263	.072914	0
S_3	0	0	0	0	0

To find the SP matrix due to S we consider the differences $S_1 - S_3$ and $S_2 - S_3$ and their variance-covariance matrix

	Variance-covariance		x^1	x^2
$S_1 - S_3$	.069099	.037263	1.1846	7.8492
$S_2 - S_3$		.072914	0.2839	2.9803

and reduce it by the square root method obtaining

.262867	.141756	4.5065	29.8600
	.229825	-1.5443	-5.4238

The SP matrix for S is obtained by multiplying the columns of this reduced matrix

$$S_{11} = (4.5065)^2 + (-1.5443)^2 = 22.69$$
$$S_{12} = (4.5065)(29.8600) + (-1.5443)(-5.4238) = 142.94$$
$$S_{22} = (29.8600)^2 + (-5.4238)^2 = 921.04$$

There is no simple (direct) method of obtaining the SP matrix for factors other than the last. Finally for F, following the same method as for S we have

	Estimates		*Variance-covariance matrix*			
	x^1	x^2				
F_1	15.5731	13.3855	.115388	.038515	.030808	.030509
F_2	17.4614	13.2030		.077292	.034021	.032269
F_3	20.8049	19.1705			.074690	.033494
F_4	11.3033	8.2452				.071393

From this the differences and variances—covariances are found

	Variance-covariance matrix			*Estimates*	
				x^1	x^2
F_1-F_4	.113582	.041040	.041108	4.2698	5.1403
F_2-F_4		.084147	.040551	6.1581	4.9578
F_3-F_4			.079095	9.5016	10.9253

By reducing this (by square root or Gauss-Doolittle) the SP matrix for F is found as above (for S)

$$S_{11} = 2123.91, \; S_{12} = 2048.81, \; S_{22} = 4094.10$$

The tests of overall significance using the SP matrices for main effects and the individual differences of the levels of each factor are carried out in the usual way.

5. Multiple classification

The treatment of four-way classification is similar to that of three-way. The general methods are as follows.

The highest order interaction is broken into components of single degrees of freedom and estimated by substituting the cell means. The variance-covariance matrix of these estimates are obtained by combining the reciprocals of the cell numbers. Then the computational technique of Table 6 is employed.

To compute any other interaction say ABC when there are other factors, the components of ABC and their variance-covariance matrices are found for each combination of the other factors as in Table 9 for the evaluation of $F \times S$. The computational scheme suggested for Table 9 yields the necessary SP matrix. This method is quite general for interaction of any order under the only condition that the higher order interactions using these factors are non-existent.

The computation of main effects follows exactly the same lines as in Table 11.

When some of the factors are at two levels a similar method can be followed for the computation of interactions. For instance if AB has to be computed when B is at two levels, first a two way table may be prepared, one way containing the levels of A and the other way the combinations of the other factors whatever their number may be. In each cell the difference in the mean values of the two levels of B is recorded together with the harmonic mean of the sample numbers for the two means. This table is treated in the same way as Table 3 or 8. The 'main effect A eliminating the other combinations' from this table provides the SP matrix for the interaction AB. Similar devices may easily be thought out on the basis of numerous computational techniques considered in this paper.

Recently at the request of Dr. J. Schull of the University of Michigan the author had the occasion to use the techniques developed in this paper for data in a four-way classification with unequal numbers in cells using four characters.

APPENDIX

The computational aspects of the general theory of least squares

In two earlier papers (Rao 1945, 1946), the author discussed a general theory of least squares when there are no restrictions on the number of unknown parameters, the rank of the matrix of the observational equations and the number of observations. In this general situation the possibilities are that not all linear parametric functions are estimable (in the sense of existence of an unbiased estimate) and the matrix of normal equations becomes singular creating some computational difficulties. But it was shown that the normal equations always admit solutions and it is enough for purposes of estimation and obtaining the least squares to have just one solution out of a possible multiplicity. The computational technique of obtaining a solution is not fully discussed. The practice has been to add to the normal equations some consistent equations and obtain a solution. The consistent equations are conveniently chosen to simplify the solution. In a general situation involving the fitting of constants by a numerical solution the problem may not be simplified that way. Also there is the question of finding the variances and covariances of the estimates for which a simple technique was provided in the general theory but the computational side needed some attention. In this appendix these computational problems have been mechanized in such a way that starting with normal equations and following the simple procedure of successive elimination one obtains a solution of the normal equations and the expressions needed to build up the variances and covariances of estimates. With this gap filled on the computational side a complete treatment of the general theory of least squares is made available.

Pivotal condensation of a semi-definite matrix: The first complication in the general theory of the least squares, as noted above, is caused by the singularity of the matrix of normal equations. When it is singular it is positive semi-definite, otherwise positive definite. The methods given below are applicable to all cases.

Pivotal condensation is a method by which a matrix A can be reduced to a triangular form. The first row is chosen and using the first element as the pivot the first column is swept out by row multiplication and addition. If the first element is zero then all elements in the column are zero and it is already in swept out form. Now from the reduced matrix the second row is chosen and using the diagonal element the elements below in the second column are swept out and the process is continued. If at any stage the diagonal element is zero, the entire row and column will be zero in which case it is already in the swept out form and the next row is chosen. The rows, chosen in the successive operations, written one after the other form a matrix where all the elements below the diagonal are necessarily zero.

Any operation on the rows is equivalent to multiplication by an elementary and hence a nonsingular matrix and therefore it follows that there exists a matrix (product of elementary matrices) C such that for the given matrix A the matrix CA has a triangular form. Let Δ be the diagonal matrix containing the diagonal elements of CA. If the rank of A is r and order m then $(m-r)$ diagonal elements in Δ and the corresponding rows in CA will all be zero. We will be considering only symmetrical matrices in which case it is easy to show that $CAC' = \Delta$.

When A has the full rank m, all the diagonal elements of CA are non-zero in which case by dividing each row of CA by the corresponding diagonal element a new matrix with unity as diagonal elements is obtained. This matrix can be represented by the product $\Delta^{-1}CA$ where Δ^{-1} is the inverse of Δ. The *Gauss-Doolittle method* of reducing equations is intended to obtain the matrix $\Delta^{-1}CA$ which has a triangular form with unity as diagonal elements to facilitate back solution. Since Δ has all positive elements each row of CA can be divided by the square root of the corresponding diagonal element to obtain the matrix $\Delta^{-\frac{1}{2}}CA$ which is the the reduced matrix obtained by the *square root method*. So both these methods give essentially the same triangular matrix except that the rows of one are multiples of the other. Given one the other can be derived by row multiplication with constants.

When the rank of A is not full some elements of Δ will be zero. Let us define by ∇ a diagonal matrix with a diagonal element being zero if the corresponding element in Δ is zero or the reciprocal if non-zero. It is evident that $\nabla = \Delta^{-1}$ when the latter exists. More generally we have

$$\nabla CA \qquad \dots \quad (A.1)$$

in a Gauss-Doolittle form and

$$\nabla^{\frac{1}{2}}CA \qquad \dots \quad (A.2)$$

in a square root form obtainable by the same method of computation as for non-singular matrices except *when a zero row is encountered in the reduction process it is left as it is.*

To check up the condition of square root we find

$$(AC'\nabla^{\frac{1}{2}})(\nabla^{\frac{1}{2}}CA) = (C^{-1}\Delta\nabla^{\frac{1}{2}})(\nabla^{\frac{1}{2}}\Delta C^{-1'})$$

$$= C^{-1}\Delta\nabla\Delta C^{-1'} = C^{-1}\Delta C^{-1'} = A \qquad \dots \quad (A.3)$$

since $\quad CA = \Delta C^{-1'}$.

This shows that a positive semi-definite matrix admits a square root which is otherwise evident.

Solution of the normal equations: In the theory of least squares we have normal equations*

$$a_{11}t_1+\ldots+a_{1m}t_m = Q_1$$
$$\cdot \quad \cdots \quad \cdot \quad \cdot$$
$$a_{m1}t_1+\ldots+a_{mm}t_m = Q_m$$

which in matrix notation could be written

$$AT = Q \qquad \ldots \quad (A.4)$$

where T and Q are column vectors.

$$AT = Q \longleftrightarrow CAT = CQ = \Delta C^{-1'}T \qquad \text{since } CA = \Delta C^{-1'}.$$

This is obtained by pivotal condensation with the column Q adjoined to matrix A. The equation

$$\Delta C^{-1'}T = CQ \qquad \ldots \quad (A.5)$$

is satisfied, because of the consistency of the normal equations, by

$$T = C'\nabla CQ \qquad \ldots \quad (A.6)$$

obtained by formally treating ∇ as the inverse of Δ and pre-multiplying both sides of (A.5) by $C'\nabla$ or pre-multiplying (A.4) by $C'\nabla C$ which so to say behaves as the inverse of $A = C^{-1}\Delta C^{-1'}$. The matrix of the equation (A.5) is triangular (has the form $\nabla CAT = \nabla CQ$ in the Gauss-Doolittle scheme and $\nabla^{\frac{1}{2}}CAT = \nabla^{\frac{1}{2}}CQ$ in the square root) so that a solution for $t_1,\ldots, t_m$ can be obtained by back solution starting with t_m.

Least squares: The expression for least sum of squares is

$$\Sigma t_iQ_i = T'Q = Q'C'\nabla CQ = (CQ)'(\nabla CQ) = (\nabla^{\frac{1}{2}}CQ)'(\nabla^{\frac{1}{2}}CQ) \qquad \ldots \quad (A.7)$$

which depends only on the matrices obtained by reducing the normal equations either by Gauss-Doolittle or square root. The following are illustrative examples.

PIVOTAL CONDENSATION WITH AN ADJOINED COLUMN

example A_1

Gauss-Doolittle technique

row	t_1	t_2	t_3	q
1	4	8	12	16
2	8	16	24	32
3	12	24	48	72
1.0	4	8	12	16
1.1	1	2	3	4
2.0		0	0	0
2.1		0	0	0
3.0			12	24
3.1			1	2
matrix		CA & ∇CA		CQ & ∇CQ

example A_2

square root

row	t_1	t_2	t_3	q
1	4	8	12	16
2	8	16	24	32
3	12	24	48	72
1.1	2	4	6	8
2.1		0	0	0
3.1			$\sqrt{12}$	$24/\sqrt{12}$
matrix		$\nabla^{\frac{1}{2}}CA$		$\nabla^{\frac{1}{2}}CQ$

*In the body of the paper, in Section 2, the coefficients of the matrix of normal equations are denoted by g instead of a which are used in the observational equations. This should not cause any confusion.

The normal equations are in the rows 1, 2, 3. We define the product of two columns $(x).(y)$ in the Gauss-Doolittle scheme as the sum of products of elements one chosen from row (i.0) from column (x) (or y) and another from row (i.1) from column (y) (or x), the summation being over all $i = 1, 2, 3$ (in this case). For the square root method $(x).(y)$ is simply the sum of products of elements in columns (x) and (y) taken from the rows (i.1), $i = 1, 2, \ldots$; With this definition we can check up in either scheme (A_1 or A_2)

$$\begin{array}{lll} (t_1).(t_1) = a_{11} & (t_1).(t_2) = a_{12} & (t_1).(t_3) = a_{13} \\ & (t_2).(t_2) = a_{22} & (t_2).(t_3) = a_{23} \\ & & (t_3).(t_3) = a_{33} \end{array}$$

which provide the theoretical basis of the method of successively building up the elements of the reduced matrices starting with the matrix (a_{ij}). The matrix in the the rows (1.0), (2.0), (3.0) in example (A_1) contain the reduced form CA of A and CQ. The rows (1.1), (2.1), (3.1) in A_1 give the matrices ∇CA, ∇CQ while in A_2 give $\nabla^{\frac{1}{2}}CA$, $\nabla^{\frac{1}{2}}CQ$; thus each row differs in the two methods only in the multiplying constant.

Back solution: From the last row (3.1), in either scheme, the value of $t_3 = 2$ is obtained. The value of t_i corresponding to zero pivotal elements may be taken as zero (or any value). Let us choose $t_2 = 0$ to correspond to solution (A.6). Substituting the values of t_3 and t_2 in (1.1) the value of $t_1 = -2$ is obtained. Finally these values may be substituted in the equations (1, 2, 3) for a check.

Least squares: The expression (A.7) needed for least squares is

$$T'Q = (CQ)'(\nabla CQ) = (\nabla^{\frac{1}{2}}CQ)'(\nabla^{\frac{1}{2}}CQ)$$

which in the notation of the above computational schemes can be written as column multiplications

$$\Sigma t_iQ_i = (q).(q) = 16\times 4+0\times 0+24\times 2 = 72 \quad \text{(Example } A_1)$$

$$(q).(q) = \ 8^2+0^2+(24/\sqrt{12})^2 = 72 \quad \text{(Example } A_2).$$

It may be observed that $\Sigma t_iQ_i = Q'A^{-1}Q$ when the rank of A is full so that the same computational technique as above is available for the evaluation of quadratic forms like $Q'A^{-1}Q$ given the matrices A and Q.

Extension to analysis of dispersion: For purposes of analysis of dispersion we have to solve several sets of normal equations with the same left hand side matrix. If Q is the matrix of right hand side elements of the equations

$$a_{11}t_1+\ldots+a_{1m}t_m = Q_1^1, Q_1^2, \ldots, Q_1^k$$

$$\cdot \quad \cdots \quad \cdot \qquad \cdot \quad \cdot \quad \cdots \quad \cdot$$

$$a_{m1}t_1+\ldots+a_{mm}t_m = Q_m^1, Q_m^2, \ldots, Q_m^k$$

and T is the matrix of solutions

$$\begin{pmatrix} t_1^1 & t_1^2 & \dots & t_1^k \\ \cdot & \cdot & \dots & \cdot \\ t_m^1 & t_m^2 & \dots & t_m^k \end{pmatrix}$$

then in matrix notation

$$AT = Q$$

and the dispersion matrix required is $T'Q$ which can be computed by first reducing the matrix A with Q adjoined to it. The (r,s)th element of $T'Q$ is $\Sigma t_i^r Q_i^s$.

PIVOTAL CONDENSATION WITH AN ADJOINED MATRIX

example A_3

Gauss-Doolittle

row	t_1	t_2	t_3	q^1	q^2
1	4	8	12	16	32
2	8	61	24	32	64
3	12	24	48	72	100
1.0	4	8	12	16	32
1.1	1	2	3	4	8
2.0		0	0	0	0
2.1		0	0	0	0
3.0			12	24	4
3.1			1	2	4/12

example A_4

square root

row	t_1	t_2	t_3	q^1	q^2
1	4	8	12	16	32
2	8	16	24	32	64
3	12	24	48	72	100
1.1	2	4	6	8	16
2.1		0	0	0	0
3.1			$\sqrt{12}$	$24/\sqrt{12}$	$4/\sqrt{12}$

scheme A_3 scheme A_4

$$\Sigma t_i^1 Q_i^1 = (q^1).(q^1) = 16\times 4+0\times 0+24\times 2 = 112 = 8^2+0^2+(24/\sqrt{12})^2$$

$$\Sigma t_i^1 Q_i^2 = (q^1).(q^2) = 16\times 8+0\times 0+24\times \frac{4}{12} = 72 = 8\times 16+0\times 0+\frac{24}{\sqrt{12}}\times\frac{4}{\sqrt{12}}$$

$$\Sigma t_i^2 Q_i^2 = (q^2).(q^2) = 32\times 8+0\times 0+4\times \frac{4}{12} = 200\frac{1}{3} = 16^2+0^2+\left(\frac{4}{\sqrt{12}}\right)^2.$$

Pseudo inverse of a singular matrix and variance-covariances of estimates: It may be observed that the solution (A.6) of the normal equations provides the estimates of $\tau_1, \dots, \tau_m$ the individual unknown parameters only when each parameter is estimable. This is so when the rank of the normal equations is full. But the estimate of any estimable parametric function $p_1\tau_1+\dots+p_m\tau_m$ can be obtained by substituting for τ_i any solution of the normal equations (Rao, 1945, 1946). Thus the estimate is using (A.6)

$$\Sigma p_i t_i = P'T = P'C'\nabla CQ$$

which has the variance

$$(P'C'\nabla CP)\sigma^2 \qquad \ldots \quad (A.8)$$

using the formula given in Rao (1945, 1952). The expression (A.8) suggests that the matrix $(C'\nabla C)\sigma^2$ behaves like the variance-covariance matrix of the formal estimates $t_1, \ldots, t_m$. Thus once a solution $t_1, \ldots, t_m$ of the normal equations and the *pseudo inverse* of A, the matrix $C'\nabla C = (e_{ij})$ are obtained, estimates of parametric functions and their variances-covariances can can be obtained in a formal way. In ordinary notation the estimate of $p_1\tau_1 + \ldots + p_m\tau_m$ is $p_1t_1 + \ldots + p_mt_m$ and its variance is

$$\Sigma\Sigma p_ip_j \operatorname{cov}(t_it_j) = \Sigma\Sigma p_ip_je_{ij}\sigma^2.$$

Pseudo inverse by back solution: This can be obtained from the reduced matrices in examples A_1 or A_2 by a series of back solutions or directly if a unit matrix is adjoined to A for reduction. When a row with a zero pivotal element is encountered, the derived row is replaced by all zeroes in the Gauss-Doolittle scheme. In the square root scheme the entire row will be zero if the pivotal element is zero.

PIVOTAL CONDENSATION WITH AN ADJOINED UNIT MATRIX

example A_3
Gauss-Doolittle

row	t_1	t_2	t_3	q	t'_1	t'_2	t'_3
1	4	8	12	16	1	.	.
2	8	16	24	32	.	1	.
3	12	24	48	72	.	.	1
1.0	4	8	12	16	1	.	.
1.1	1	2	3	4	$\underline{1/4}$	.	.
2.0		0	0	0	−2	1	.
2.1		0	0	0	0	$\underline{0}$	0
3.0			12	24	−3	0	1
3.1			1	2	−1/4	0	$\underline{1/12}$

example A_4
square root

row	t_1	t_2	t_3	q	t'_1	t'_2	t'_3
1	4	8	12	16	1	.	.
2	8	16	24	32	.	1	.
3	12	24	48	72	.	.	1
1.1	2	4	6	8	$\underline{1/2}$	.	.
2.1		0	0	0	0	$\underline{0}$	0
2.2			$\sqrt{12}$	$24/\sqrt{12}$	$-3/\sqrt{12}$		$\underline{1/\sqrt{12}}$

(a) To obtain the pseudo inverse by a series of back solutions it is not necessary (see Dwyer, 1951, p.170) to reduce the entire unit matrix. The elements underlined, which are the reciprocals of non-zero pivotal elements, are all what is needed. Let us solve the equations with the elements under (t'_3) as the right hand side elements. The back solution gives

$$t_{33} = \frac{1}{12}, \quad t_{23} = 0, \quad t_{13} = -\frac{1}{4}$$

Next let us solve the equations for (t'_2) and observe that the value of t_3, t_{32} is same as t_{23} because of symmetry. Now back solution for t_{22} and t_{12} gives

$$(t_{32} = 0), \quad t_{22} = 0, \quad t_{12} = 0.$$

Now for the equation under (t'_1), the solutions t_{31}, t_{21} are known by symmetry. Therefore t_{11} is obtained by back solution

$$(t_{31} = -\frac{1}{4}\,,\quad t_{21} = 0),\quad t_{11} = 1.$$

The complete pseudo inverse matrix, omitting the diagonal elements below, is

$$(e_{ij}) = \begin{pmatrix} 1 & 0 & -1/4 \\ & 0 & 0 \\ & & 1/12 \end{pmatrix}$$

(b) But if the entire unit matrix is reduced we can obtain the inverse elements and the solutions directly without back solution in any order we like. To obtain the solutions

scheme A_5 scheme A_6

$$t_1 = (t'_1).(q) = 16\left(\frac{1}{4}\right)+0(0)+24\left(-\frac{1}{4}\right) = -2 = 8\left(\frac{1}{2}\right)+0(0)+\frac{24}{\sqrt{12}}\left(-\frac{3}{\sqrt{12}}\right)$$

$$t_2 = (t'_2).(q) = 16\,(0)\,+0(0)+24\,(\,0\,) = \quad 0 = 8\,(0)\,+0(0)+\frac{24}{\sqrt{12}}\,(0)$$

$$t_3 = (t'_3).(q) = 16\,(0)\,+0(0)+24\left(\frac{1}{12}\right) = \quad 2 = 8\,(0)\,+0(0)+\frac{24}{\sqrt{12}}\left(\frac{1}{\sqrt{12}}\right)$$

The inverse elements are

scheme A_5 scheme A_6

$$e_{11} = (t'_1).(t'_1) = 1\left(\frac{1}{4}\right)+(-2)(0)+(-3)\left(-\frac{1}{4}\right) = 1 = \left(\frac{1}{2}\right)^2+(0)^2+\left(\frac{-3}{\sqrt{12}}\right)^2$$

$$e_{12} = (t'_1).(t'_2) = 1\,(0)\,+(-2)(0)+(-3)\,(\,0\,) = 0 = \left(\frac{1}{2}\right)(0)+(0)(0)+\left(\frac{-3}{\sqrt{12}}\right)(0)$$

etc. etc. etc.

This method is particularly useful in problems where the entire inverse is not needed as in the illustrations of Section 4.

It appears that singular matrices create no difficulty in the computations necessary to set up analysis of variance and dispersion tables. In fact they need not be recognized beforehand. They will, however, be discovered during the process of computation. If it is predictable as to which rows are likely to become zero it may save trouble to omit those rows and columns to start with. The constants corresponding to them will have zero values for solution and so also their variances and covariances with the other constants. In the illustrations of this paper the rows and columns have been omitted in some cases by prior examination. But this is unnecessary because their retention causes no trouble. May be it pays to reserve our judgement because there is a possibility of wrong judgement. What is more important, non-omission provides a further check on the accuracy of computations because these rows at some stage of reduction should vanish (up to the number of significant figures provided).

REFERENCES

DWYER, P. S. (1951): *Linear computations*. John Wiley & Sons, New York.

FISHER, R. A. (1939): The sampling distribution of some statistics obtained from non-linear regression. *Ann. Eugen.*, 9, 238.

GOULDEN, C. H. (1952): *Methods of Statistical Analysis*. John Wiley & Sons, New York.

KENDALL, M. G. (1946): *Advanced Theory of Statistics*, Charles Griffin & Co. Ltd., London.

RAO, C. R. (1945): Generalisation of Markoff's theorem and tests of linear hypotheses. *Sankhyā*, 7, 9.

——— (1946): On the linear combination of observations and the general theory of least squares. *Sankhyā*, 7, 237.

——— (1952): *Advanced Statistical Methods in Biometric Research*. John Wiley & Sons, New York.

Paper received : May, 1955.