Biometrika (1971), 58, 3, *p.* 545
Printed in Great Britain

Recovery of inter-block information when block sizes are unequal

BY H. D. PATTERSON AND R. THOMPSON
University of Edinburgh

SUMMARY

A method is proposed for estimating intra-block and inter-block weights in the analysis of incomplete block designs with block sizes not necessarily equal. The method consists of maximizing the likelihood, not of all the data, but of a set of selected error contrasts. When block sizes are equal results are identical with those obtained by the method of Nelder (1968) for generally balanced designs. Although mainly concerned with incomplete block designs the paper also gives in outline an extension of the modified maximum likelihood procedure to designs with a more complicated block structure.

1. INTRODUCTION

In this paper we consider the estimation of weights to be used in the recovery of inter-block information in incomplete block designs with possibly unequal block sizes. The problem can also be thought of as one of estimating constants and components of variance from data arranged in a general two-way classification when the effects of one classification are regarded as fixed and the effects of the second classification are regarded as random.

Nelder (1968) described the efficient estimation of weights in generally balanced designs, in which the blocks are usually, although not always, of equal size. Lack of balance resulting from unequal block sizes is, however, common in some experimental work, for example in animal breeding experiments. The maximum likelihood procedure described by Hartley & Rao (1967) can be used but does not give the same estimates as Nelder's method in the balanced case.

As will be shown, the two methods in effect use the same weighted sums of squares of residuals but assign different expectations. In the maximum likelihood approach, expectations are taken over a conditional distribution with the treatment effects fixed at their estimated values. In contrast Nelder uses unconditional expectations.

The difference between the two methods is analogous to the well-known difference between two methods of estimating the variance σ^2 of a normal distribution, given a random sample of n values. Both methods use the same total sum of squares of deviations. But whereas one method equates the sum of squares to $(n-1)\sigma^2$ the other equates the sum of squares to $n\sigma^2$. The former method gives an unbiased estimate of σ^2; the latter maximizes the likelihood of the sample.

Another method for unbalanced designs has been described by Cunningham & Henderson (1968). When corrected as described by Thompson (1969) this method allows for errors in the estimation of treatment effects but in general the estimates are not efficient.

The method proposed in the present paper is a modified maximum likelihood procedure, more efficient than the Cunningham and Henderson method and giving the same results as Nelder's method in the analysis of balanced designs. The contrasts among yields are divided into two sets: (i) contrasts between treatment totals; and (ii) contrasts with zero expectation, i.e. error contrasts. The method consists of maximizing the joint likelihood of all possible contrasts in set (ii). Contrasts in set (i) are excluded from the likelihood function on the grounds that, as long as treatment effects are regarded as unknown, fixed, as opposed to random, and without restraints, no contrast in set (i) can provide any information on error.

The modified maximum likelihood method was considered by Patterson (1964) in a components-of-variance problem arising in the analysis of rotation experiments. This paper was, however, primarily concerned with evaluating the efficiencies of simpler methods and gave no details beyond formulae for asymptotic variances.

2. THE MODEL

We suppose that the incomplete block design has t treatments and n units, plots, in b blocks of possibly unequal size and that the $n \times 1$ vector of yields, $\mathbf{y}$, can be represented by the model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \boldsymbol{\epsilon}, \tag{1}$$

where $\mathbf{X}$ is an $n \times t$ matrix of rank t determined by the allocation of treatments to units, $\boldsymbol{\alpha}$ is a $t \times 1$ vector of treatment parameters and $\boldsymbol{\epsilon}$ is a random variable normally distributed with mean zero and variance given by

$$\mathbf{V} = \mathbf{H}\sigma^2, \quad \mathbf{H} = \mathbf{Z}\boldsymbol{\Gamma}\mathbf{Z}' + \mathbf{I}. \tag{2}$$

Further $\boldsymbol{\Gamma} = \gamma\mathbf{I}$, $\mathbf{Z}$ is an $n \times b$ matrix with elements Z_{ij} equal to 1 when unit i is in block j ($i = 1, \ldots, n$; $j = 1, \ldots, b$) and equal to 0 otherwise, and γ and σ^2 are unknown scalars. A model with more general $\mathbf{Z}$ and $\boldsymbol{\Gamma}$ is considered in § 10.

The problem is to estimate $\boldsymbol{\alpha}$, γ and σ^2. Sometimes reparameterization may be possible. Provided γ is not negative, $\gamma\sigma^2$ and σ^2 can be regarded as components of variance σ_b^2 and σ_e^2, say. When blocks sizes are all equal to k we may require to estimate functions

$$V_1 = (k\gamma + 1)\,\sigma^2, \quad V_2 = \sigma^2.$$

This is, in fact, what Nelder (1968) does.

The matrix $\mathbf{H}$ is essentially nonsingular. The inverse $\mathbf{H}^{-1}$ can be written in the form

$$\mathbf{H}^{-1} = \mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z} + \boldsymbol{\Gamma}^{-1})^{-1}\mathbf{Z}'. \tag{3}$$

Thus $\mathbf{H}^{-1}$ exists if $(\mathbf{Z}'\mathbf{Z} + \boldsymbol{\Gamma}^{-1})^{-1}$ exists; $\mathbf{Z}'\mathbf{Z}$ is diagonal with elements k_j, where k_j is the number of plots in block j. Clearly $\mathbf{Z}'\mathbf{Z} + \boldsymbol{\Gamma}^{-1}$ is singular if $\gamma = -k_j^{-1}$ for some j. But this would imply that the variance of the mean for block j is zero. A smaller value of γ implies a negative variance. We can, therefore, reasonably impose the condition that $\gamma > -1/k_{\max}$, where $k_{\max}$ is the largest number of plots in a single block.

Other conditions must also be satisfied for estimation of γ and σ^2 to be possible. These conditions will be considered in § 6.

Error model (2) also implies that the correlation between two units in the same block is independent of block size. This model is often used in animal experiments with blocks consisting of genetically related animals; usually the relationships within a block can be assumed to be independent of the number of animals in the block. In some other applications, for example field plot experiments, it might be more appropriate to specify a correlation that varies from block to block with the largest blocks showing the smallest correlations.

Estimates of $\boldsymbol{\alpha}$, γ and σ^2 will be denoted by $\hat{\boldsymbol{\alpha}}$, $\hat{\gamma}$ and $\hat{\sigma}^2$. We will also use the circumflex to denote functions of $\hat{\gamma}$, for example $\hat{\boldsymbol{\Gamma}} = \hat{\gamma}\mathbf{I}$ and $\hat{\mathbf{H}} = \mathbf{Z}\hat{\boldsymbol{\Gamma}}\mathbf{Z}' + \mathbf{I}$.

3. MODIFIED MAXIMUM LIKELIHOOD METHOD

The logarithm of the likelihood function of $\mathbf{y}$ is given by

$$L = \text{const} - \tfrac{1}{2}\log|\mathbf{H}| - \tfrac{1}{2}n\log\sigma^2 - \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})'\mathbf{H}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}). \tag{4}$$

Hartley & Rao (1967) estimate $\boldsymbol{\alpha}$, γ and σ^2 by maximizing L. In the present paper we divide the data into two parts, with separate logarithmic likelihoods L' and L'', estimate γ and σ^2 by maximizing L' and estimate $\boldsymbol{\alpha}$ by maximizing L''.

The two parts can be represented by transformed yields $\mathbf{Sy}$ and $\mathbf{Qy}$ with the following properties.

(i) The matrix $\mathbf{S}$ is of rank $n - t$ and $\mathbf{Q}$ is a matrix of rank t.

(ii) The two parts are statistically independent, i.e. $\mathrm{cov}(\mathbf{Sy}, \mathbf{Qy}) = 0$. This condition is met if

$$\mathbf{SHQ}' = 0. \tag{5}$$

(iii) The matrix $\mathbf{S}$ is chosen so that

$$E(\mathbf{Sy}) = 0, \quad \text{i.e. } \mathbf{SX} = 0. \tag{6}$$

(iv) The matrix $\mathbf{QX}$ is of rank t, so that every linear function of the elements of $\mathbf{Qy}$ estimates a linear function of the elements of $\boldsymbol{\alpha}$.

It follows from (i) and (ii) that the likelihood of $\mathbf{y}$ is the product of the likelihoods of $\mathbf{Sy}$ and $\mathbf{Qy}$, i.e.

$$L = L' + L''. \tag{7}$$

Suitable matrices $\mathbf{S}$ and $\mathbf{Q}$ are given by

$$\mathbf{S} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}', \tag{8}$$

$$\mathbf{Q} = \mathbf{X}'\mathbf{H}^{-1}. \tag{9}$$

The matrix $\mathbf{S}$ is symmetric, idempotent, of rank $n - t$ and independent of γ. The elements of $\mathbf{Sy}$ are deviations of yields from treatment means. An estimate of γ is required for the transformation $\mathbf{Qy}$.

4. ESTIMATION OF γ AND σ^2

First, we estimate γ and σ^2 by maximizing L', the logarithmic likelihood of $\mathbf{Sy}$. The variance matrix $\mathbf{SHS}\sigma^2$ is singular but suitable expressions for L' are available in terms of either a generalized inverse $(\mathbf{SHS})^{-g}$ or the latent roots of $\mathbf{SHS}$.

In deriving these expressions we use an $n \times (n-t)$ matrix $\mathbf{P}$ whose columns are orthogonal vectors of both $\mathbf{S}$ and $\mathbf{SHS}$. As $\mathbf{S}$ is idempotent and symmetric, it can be expressed in the form $\mathbf{AA}'$, where $\mathbf{A}$ is an $n \times (n-t)$ matrix such that $\mathbf{A}'\mathbf{A} = \mathbf{I}$. Now let $\mathbf{B}$ be an orthogonal matrix such that $\mathbf{B}'\mathbf{A}'\mathbf{HAB}$ is diagonal.

The required matrix $\mathbf{P}$ is given by $\mathbf{AB}$. It has the following properties: (i) $\mathbf{P}'\mathbf{P} = \mathbf{I}$, (ii) $\mathbf{PP}' = \mathbf{S}$ and (iii) $\mathbf{P}'\mathbf{HP} = \mathrm{diag}(\xi_s)$, a diagonal matrix with elements ξ_s ($s = 1, \ldots, n-t$). We note from (i) and (ii) that $\mathbf{P}'\mathbf{S} = \mathbf{P}'$. Hence $\mathbf{P}'\mathbf{y}$ can be derived from $\mathbf{Sy}$ by the transformation $\mathbf{P}'(\mathbf{Sy})$. Also $\mathbf{Sy}$ can be derived from $\mathbf{P}'\mathbf{y}$ by the transformation $\mathbf{P}(\mathbf{P}'\mathbf{y})$. It follows that the likelihood of $\mathbf{Sy}$ is also the likelihood of $\mathbf{P}'\mathbf{y}$.

Property (iii) shows that

$$\mathbf{P'HS} = \text{diag}\,(\xi_s)\,\mathbf{P'}, \tag{10}$$

i.e. the ξ_s are the nonzero latent roots of **HS**. We can now write **SHS** in the spectral form

$$\mathbf{SHS} = \mathbf{P}\,\text{diag}\,(\xi_s)\,\mathbf{P'} \tag{11}$$

and define a generalized inverse $(\mathbf{SHS})^{-g}$ such that

$$(\mathbf{SHS})^{-g} = \mathbf{P}\,\text{diag}\,(\xi_s^{-1})\,\mathbf{P'}. \tag{12}$$

The product of $(\mathbf{SHS})^{-g}$ with **SHS** is **S**. It is also worth noting that $(\mathbf{SHS})^{-g}$ is unchanged on multiplication by **S**.

The elements of $\mathbf{P'y}$ are contrasts u_s $(s = 1, \ldots, n-t)$ with variances $\xi_s \sigma^2$. Hence the required log likelihood function is given by

$$L' = \text{const} - \tfrac{1}{2}\sum_s \log \xi_s - \tfrac{1}{2}(n-t)\log\sigma^2 - R/(2\sigma^2), \tag{13}$$

where R is the weighted sum of squares of the u_s given by

$$R = \Sigma\,(u_s^2/\xi_s) = \mathbf{y'}(\mathbf{SHS})^{-g}\,\mathbf{y}. \tag{14}$$

The estimates $\hat{\gamma}$ and $\hat{\sigma}^2$ maximizing L' are obtained by solving the equations

$$\frac{\partial L'}{\partial \gamma} = -\tfrac{1}{2}E + \frac{B}{2\sigma^2} = 0, \tag{15}$$

$$\frac{\partial L'}{\partial \sigma^2} = -\frac{1}{2}\frac{n-t}{\sigma^2} + \frac{R}{2\sigma^4} = 0, \tag{16}$$

where

$$B = \sum_s u_s^2 d_s/\xi_s^2 = \mathbf{y'}(\mathbf{SHS})^{-g}\,\mathbf{SDS}(\mathbf{SHS})^{-g}\,\mathbf{y}, \tag{17}$$

$$E = \sum_s (d_s/\xi_s) = \text{tr}\,\{(\mathbf{SHS})^{-g}\,\mathbf{SDS}\}. \tag{18}$$

Further d_s denotes $\partial\xi_s/\partial\gamma$ and **D** denotes $\partial\mathbf{H}/\partial\gamma$.

The expected values of $\partial L'/\partial\gamma$ and $\partial L'/\partial\sigma^2$ for fixed γ are both zero. Solution of equations (15) and (16) consists therefore of equating B and R to their expectations in the conditional distribution with γ fixed.

The information matrix is

$$\frac{1}{2}\begin{bmatrix} f_{11} & f_{12}/\sigma^2 \\ f_{12}/\sigma^2 & f_{22}/\sigma^4 \end{bmatrix},$$

where

$$f_{11} = \sum_s (d_s/\xi_s)^2 = \text{tr}\,\{(\mathbf{SHS})^{-g}\,\mathbf{SDS}\}^2, \tag{19}$$

$$f_{12} = E, \tag{20}$$

$$f_{22} = n - t. \tag{21}$$

So far the results are appropriate when **H** is any symmetric $n \times n$ matrix of rank n such that rank(**SHS**) = rank(**S**) = $n-t$. For practical purposes, however, and for comparison with other methods of estimation, we require simpler expressions for B, R, E and the elements of the information matrix. When **H** takes the particular form specified in (2) suitable expressions can be obtained by substituting the appropriate differentials **D** and d_s and a simplified form of $(\mathbf{SHS})^{-g}$.

The differential **D** is simply $\mathbf{ZZ'}$ and the d_s are given by the latent roots of $\mathbf{ZZ'S}$, or $\mathbf{SZZ'S}$. The latter can be demonstrated by substituting for **H** in equation (11) and rearranging to give

$$\mathbf{SZZ'S} = \mathbf{P}\,\text{diag}\,(\lambda_s)\,\mathbf{P'}, \tag{22}$$

where the λ_s are such that

$$\xi_s = \lambda_s \gamma + 1. \tag{23}$$

Hence $d_s = \lambda_s$.

We now express $(\mathbf{SHS})^{-g}$ in terms of a $b \times b$ matrix. First, we note that the factors of $\mathbf{Z'SZ\Gamma Z'S}$ include both the $b \times b$ matrix $\mathbf{Z'SZ}$ and the $n \times n$ matrix $\mathbf{SZ\Gamma Z'S} = \mathbf{S(H-I)S}$. Hence

$$(\mathbf{Z'SZ})\,\mathbf{\Gamma Z'S} + \mathbf{Z'S} = \mathbf{Z'S}(\mathbf{SHS}),$$

so that

$$\mathbf{\Gamma Z'}(\mathbf{SHS})^{-g} = \mathbf{W}^{-1}\mathbf{Z'S}, \tag{24}$$

where

$$\mathbf{W} = \mathbf{Z'SZ} + \mathbf{\Gamma}^{-1}. \tag{25}$$

Premultiplying both sides of (24) by **SZ**, we obtain

$$(\mathbf{SHS} - \mathbf{S})\,(\mathbf{SHS})^{-g} = \mathbf{SZW}^{-1}\mathbf{Z'S},$$

so that

$$(\mathbf{SHS})^{-g} = \mathbf{S} - \mathbf{SZW}^{-1}\mathbf{Z'S}. \tag{26}$$

The quantities B, R, E and f_{11} can now be expressed in terms of **W**, a vector

$$\boldsymbol{\beta} = \mathbf{\Gamma Z'}(\mathbf{SHS})^{-g}\,\mathbf{y} = \mathbf{W}^{-1}\mathbf{Z'Sy} \tag{27}$$

and a matrix

$$\mathbf{U} = \mathbf{Z'}(\mathbf{SHS})^{-g}\mathbf{Z} = \mathbf{\Gamma}^{-1} - \mathbf{\Gamma}^{-1}\mathbf{W}^{-1}\mathbf{\Gamma}^{-1}. \tag{28}$$

The required expressions for B and R can be obtained by direct substitution. They are

$$B = \boldsymbol{\beta}'\boldsymbol{\beta}/\gamma^2, \tag{29}$$

$$R = \mathbf{y'Sy} - \mathbf{y'SZ}\boldsymbol{\beta}. \tag{30}$$

Also, since $\text{tr}\,\{(\mathbf{SHS})^{-g}\,\mathbf{ZZ'S}\}$ is equivalent to $\text{tr}\,\{\mathbf{Z'}(\mathbf{SHS})^{-g}\,\mathbf{Z}\}$, E is given by

$$E = f_{12} = \text{tr}\,(\mathbf{U}). \tag{31}$$

Similarly,

$$f_{11} = \text{tr}\,(\mathbf{U}^2). \tag{32}$$

In practice, of course, we use $\hat{\mathbf{H}}$, $\hat{B}$, $\hat{\boldsymbol{\beta}}$, $\hat{R}$, etc., given by the expressions for **H**, B, $\boldsymbol{\beta}$, R, etc., but with γ replaced by $\hat{\gamma}$.

These results depend on **W** being nonsingular; in fact **W** is always nonsingular under the conditions imposed on **H** in § 2 as the largest λ_s is necessarily not larger than $k_{\max}$.

5. Estimation of $\boldsymbol{\alpha}$

The estimate of $\boldsymbol{\alpha}$ is obtained by maximizing

$$L'' = \text{const} - \tfrac{1}{2}\log|\mathbf{X'H}^{-1}\mathbf{X}| - \tfrac{1}{2}t\log\sigma^2 - \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})'\,\mathbf{H}^{-1}\mathbf{X}(\mathbf{X'H}^{-1}\mathbf{X})^{-1}\mathbf{X'H}(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}). \tag{33}$$

The estimate is given by

$$\hat{\boldsymbol{\alpha}} = (\mathbf{X'}\hat{\mathbf{H}}^{-1}\mathbf{X})^{-1}\mathbf{X'}\hat{\mathbf{H}}^{-1}\mathbf{y}. \tag{34}$$

An equivalent procedure is to minimize the weighted sum of squares

$$(\mathbf{y}-\mathbf{X}\boldsymbol{\alpha})'\,\hat{\mathbf{H}}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\alpha})$$

with fixed $\hat{\gamma}$.

6. Practical solution of equations (15) and (16)

Equations (15) and (16) can be solved by Fisher's iterative method. We start with a preliminary estimate, $\hat{\gamma}_0$ say, of γ. Substitute this estimate for γ in (25) to give $\hat{\mathbf{W}}$, and hence determine $\hat{\boldsymbol{\beta}}$, $\hat{\mathbf{U}}$, $\hat{B}$, $\hat{R}$, $\hat{f}_{12}$ and $\hat{f}_{11}$ equations from (27) to (32). Then an approximation to $\hat{\sigma}^2$ is given by

$$\hat{\sigma}^2 = \hat{f}^{12}\hat{B}+\hat{f}^{22}\hat{R}, \tag{35}$$

and a closer approximation to $\hat{\gamma}$ is given by

$$\hat{\gamma} = \hat{\gamma}_0+(\hat{f}^{11}\hat{B}+\hat{f}^{12}\hat{R})/\hat{\sigma}^2. \tag{36}$$

In these expressions $\hat{f}^{11}$, $\hat{f}^{12}$ and $\hat{f}^{22}$ are the elements of the inverse matrix $\hat{\mathbf{F}}^{-1}$, where

$$\hat{\mathbf{F}} = \begin{bmatrix}\hat{f}_{11} & \hat{f}_{12}\\ \hat{f}_{12} & \hat{f}_{22}\end{bmatrix}.$$

If $\hat{\gamma}_0$ is small it may be more accurate numerically not to calculate $\hat{\boldsymbol{\beta}}$ explicitly, but to use instead

$$\hat{\boldsymbol{\beta}}/\hat{\gamma} = (\hat{\gamma}_0\mathbf{Z}'\mathbf{S}\mathbf{Z}+\mathbf{I})^{-1}\mathbf{Z}'\mathbf{S}\mathbf{y}.$$

A preliminary estimate of σ^2 is not required. Also $\boldsymbol{\alpha}$ need not be estimated until $\hat{\gamma}$ has been determined. A convenient expression for $\hat{\boldsymbol{\alpha}}$ in terms of $\hat{\gamma}$ can be obtained by noting that $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ in (34) and (27) satisfy the equations

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\alpha}}+\mathbf{X}'\mathbf{Z}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}, \tag{37}$$

$$\mathbf{Z}'\mathbf{X}\hat{\boldsymbol{\alpha}}+(\mathbf{Z}'\mathbf{Z}+\hat{\boldsymbol{\Gamma}}^{-1})\hat{\boldsymbol{\beta}} = \mathbf{Z}'\mathbf{y}. \tag{38}$$

Hence

$$\hat{\boldsymbol{\alpha}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{y}-\mathbf{Z}\hat{\boldsymbol{\beta}}). \tag{39}$$

Equations (37) and (38) were suggested by Henderson (Henderson, Kempthorne, Searle & von Krosigk, 1959) as the basis of a practical procedure for estimating $\boldsymbol{\alpha}$ when γ is known.

When b is large an alternative procedure may be preferable, requiring inversion of a $t\times t$ matrix and a diagonal $b\times b$ matrix instead of the nondiagonal $\mathbf{W}$.

Equation (34) is used first to estimate $\boldsymbol{\alpha}$. Equation (3) provides a convenient expression for $\hat{\mathbf{H}}^{-1}$, involving inversion of a diagonal matrix only. Then

$$\hat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z}+\hat{\boldsymbol{\Gamma}}^{-1})^{-1}\mathbf{Z}'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\alpha}}), \tag{40}$$

a rearrangement of equation (38). The value of $\hat{B}$ is calculated as before and $\hat{R}$ is given by

$$\hat{R} = \mathbf{y}'\mathbf{y}-\mathbf{y}'\mathbf{X}\hat{\boldsymbol{\alpha}}-\mathbf{y}'\mathbf{Z}\hat{\boldsymbol{\beta}}. \tag{41}$$

The matrix $\hat{\mathbf{W}}^{-1}$, required in the calculation of $\hat{f}_{11}$ and $\hat{f}_{12}$, can be conveniently expressed in terms of the matrices $(\mathbf{Z}'\mathbf{Z}+\hat{\boldsymbol{\Gamma}}^{-1})^{-1}$ and $(\mathbf{X}'\hat{\mathbf{H}}^{-1}\mathbf{X})^{-1}$ already used in the estimation of $\boldsymbol{\alpha}$. The expression is

$$\hat{\mathbf{W}}^{-1} = (\mathbf{Z}'\mathbf{Z}+\hat{\boldsymbol{\Gamma}}^{-1})^{-1}+(\mathbf{Z}'\mathbf{Z}+\boldsymbol{\Gamma}^{-1})^{-1}\mathbf{Z}'\mathbf{X}(\mathbf{X}'\hat{\mathbf{H}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z}+\hat{\boldsymbol{\Gamma}}^{-1})^{-1}. \tag{42}$$

Solution of equations (15) and (16) breaks down if $\mathbf{F}$ is singular. We distinguish three cases. When $f_{22} = 0$, i.e. $n = t$, neither γ nor σ^2 can be estimated as the treatment contrasts account for all $n-1$ degrees of freedom. When $f_{11} = 0$ but $f_{22} \neq 0$, σ^2 can be estimated but not γ. This situation arises when some treatment comparison is totally confounded with every block comparison. Each column of $\mathbf{Z}$ is then given by a linear combination of the columns of $\mathbf{X}$ so that $\mathbf{SZ} = 0$ and $\mathbf{W}^{-1} = \boldsymbol{\Gamma}$, and hence $\mathbf{U} = 0$. The matrix $\mathbf{F}$ is also singular when treatment contrasts account for all intra-block degrees of freedom but not for all inter-block degrees of freedom. In this case $f_{11} = (n-t)/(\gamma+1)^2$, $f_{12} = (n-t)/(\gamma+1)$; $(\gamma+1)\,\sigma^2$ can be estimated but not the individual γ and σ^2.

A further complication is that $\mathbf{X}'\mathbf{X}$ may be singular This can be dealt with by replacing $(\mathbf{X}'\mathbf{X})^{-1}$ and $(\mathbf{X}'\mathbf{H}^{-1}\mathbf{X})^{-1}$ in (8), (34), (39) and (42) by generalized inverses and making t the rank of $\mathbf{X}$ instead of the number of treatments.

7. Relationship with other methods

We now consider the relationship between I, the method described by Hartley & Rao (1967), II, the method proposed by Cunningham & Henderson (1968) and later modified by Thompson (1969) and III, the method of the present paper. These three methods give different estimates of γ but use essentially the same methods of estimating $\boldsymbol{\alpha}$ and σ^2 once $\hat{\gamma}$ has been determined.

Hartley & Rao (1967) show that the unconditional maximum likelihood estimates are obtained by solving the following equations for $\boldsymbol{\alpha}$, σ^2 and γ:

$$\frac{1}{\sigma^2}(\mathbf{X}'\mathbf{H}^{-1}\mathbf{y}-\mathbf{X}'\mathbf{H}^{-1}\mathbf{X}\boldsymbol{\alpha}) = 0, \tag{43}$$

$$-\tfrac{1}{2}\operatorname{tr}(\mathbf{H}^{-1}\mathbf{Z}\mathbf{Z}')+\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\alpha})'\mathbf{H}^{-1}\mathbf{Z}\mathbf{Z}'\mathbf{H}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\alpha}) = 0, \tag{44}$$

$$-\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}(\mathbf{y}-\mathbf{X}\boldsymbol{\alpha})'\mathbf{H}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\alpha}) = 0. \tag{45}$$

The numerical terms $(\mathbf{y}-\mathbf{X}\boldsymbol{\alpha})'\mathbf{H}^{-1}\mathbf{Z}\mathbf{Z}'\mathbf{H}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\alpha})$ and $(\mathbf{y}-\mathbf{X}\boldsymbol{\alpha})'\mathbf{H}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\alpha})$ are equivalent to B and R. This can be demonstrated as follows. Consider the expressions for L, L' and L'' given in (4), (13) and (33). As $L = L'+L''$, we have

$$(\mathbf{y}-\mathbf{X}\boldsymbol{\alpha})'\mathbf{H}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\alpha}) = R+(\mathbf{y}-\mathbf{X}\boldsymbol{\alpha})'\mathbf{H}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{H}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{H}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\alpha}). \tag{46}$$

Substitution of the solution for $\boldsymbol{\alpha}$ given in §5, or the solution given by equation (43), shows that the second term on the right hand side of (46) is zero. Hence

$$R = (\mathbf{y}-\mathbf{X}\boldsymbol{\alpha})'\mathbf{H}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\alpha}). \tag{47}$$

Differentiation of both sides of equation (47) with respect to γ gives

$$B = (\mathbf{y}-\mathbf{X}\boldsymbol{\alpha})'\mathbf{H}^{-1}\mathbf{Z}\mathbf{Z}'\mathbf{H}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\alpha}). \tag{48}$$

Thus the Hartley–Rao method consists of equating B and R to their expected values in the conditional distribution with both γ and $\boldsymbol{\alpha}$ fixed. It differs from method III in that in the latter the expected values take account of errors in the estimation of $\boldsymbol{\alpha}$.

Method II differs from method III in that the sum of squares $\boldsymbol{\beta}'\mathbf{Z}'\mathbf{S}\mathbf{y}$ is used instead of B.

Thus $\boldsymbol{\beta}'\mathbf{Z}'\mathbf{S}\mathbf{y}$ and R are equated to their expected values in the conditional distribution with fixed γ. As in method III the estimates of γ and σ^2 are wholly derived from $\mathbf{Sy}$. Method II does not, however, maximize the likelihood of $\mathbf{Sy}$ except in special cases when $\boldsymbol{\beta}$ is proportional to $\mathbf{Z}'\mathbf{Sy}$, for example in complete block designs and in symmetric balanced incomplete block designs.

8. Equal block sizes

Nelder (1968) proposed a method for estimating stratum variances in a general class of balanced designs. We now show that application of this method to incomplete block designs with blocks of equal size k gives the same results as method III.

The method consists of equating the sums of squares of residuals,

$$(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\alpha}})'\left(\frac{\mathbf{ZZ}'}{k}-\frac{\mathbf{11}'}{n}\right)(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\alpha}}), \tag{49}$$

$$(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\alpha}})'\left(\mathbf{I}-\frac{\mathbf{ZZ}'}{k}\right)(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\alpha}}) \tag{50}$$

to their expectations, where $\mathbf{1}$ is a unit vector.

Equations (48) and (47) show that $\hat{B}$ and $\hat{R}$, the quantities used in method III, are also sums of squares of residuals. When block sizes are all equal to k the expressions for $\hat{B}$ and $\hat{R}$ simplify to

$$\hat{B} = (\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\alpha}})'\,\mathbf{ZZ}'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\alpha}})/(\hat{\gamma}k+1)^2,$$

$$\hat{R} = (\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\alpha}})'\,(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\alpha}})-\hat{\gamma}\hat{B};$$

$\mathbf{1}'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\alpha}})$, the sum of residuals, is zero. Equating $\hat{B}$ and $\hat{R}$ to their expectations is therefore equivalent to equating the expressions (49) and (50) to their expectations, i.e. the method proposed in the present paper gives the same results as Nelder's (1968) method.

Table 1. *Data used by Cunningham & Henderson* (1968)

	Treatment 1		Treatment 2		Totals
Block 1	3, 2,	5	2, 3,	5	10
2	2, 3, 5, 6, 7,	23	8, 8, 9	25	48
3	3	3	4, 4, 3, 2, 5	18	21
Totals		31		48	79

Table 2. *Estimation of* σ^2, γ *from the data of Table* 1. *Initial estimate of* $\gamma = 1$

	Cycle 1	2	3	4
B	6·3194	2·7584	2·8537	2·8505
R	42·7221	40·2145	40·2981	40·2952
$\hat{f}^{11}$	0·8155	1·8424	1·7817	1·7836
$10\hat{f}^{12}$	−0·8530	−1·2824	−1·2611	−1·2618
$10\hat{f}^{32}$	0·7142	0·7143	0·7143	0·7143
$\hat{\sigma}^2$	2·5123	2·5186	2·5185	2·5185
$\hat{\gamma}$	1·6006	1·5708	1·5718	1·5718

9. Example

We have applied the new method, method III, to the data of the example discussed by Cunningham & Henderson (1968) and Thompson (1969). The data are in Table 1. Table 2 gives the results of four cycles of the iterative procedure, starting with $\hat{\gamma}_0 = 1$. Convergence is rapid. Table 3 compares the results with those given by methods I and II and by the analysis of variance method described by Cunningham & Henderson (1968).

Table 3. *Comparison of estimates of* σ^2 *and* γ

Method	Estimate of σ^2	Estimate of γ
Analysis of variance	2·5237	1·7479
Method I	2·3518	1·0652
Method II	2·4822	1·8028
Method III	2·5185	1·5718

10. More general results

Hartley & Rao (1967) extended the unconditional maximum likelihood method to a general class of designs with c block factors. The class includes for example split-plot designs and row-and-column designs. With these designs the model given by equation (1) still applies but the variance matrix of $\boldsymbol{\epsilon}$ is now $\mathbf{H}\sigma^2$, where

$$\mathbf{H} = \left(\mathbf{I}+\sum_{p=1}^{c}\mathbf{Z}_p\mathbf{Z}_p'\gamma_p\right). \tag{51}$$

We have to estimate $c+1$ parameters, σ^2 and γ_p $(p = 1, \ldots, c)$.

Each plot is at exactly one level of each block factor. For example, if a design is arranged in rows and columns each plot is in exactly one row and one column. Element (i,j) of $\mathbf{Z}_p$ is 1 when plot i is at level j of block factor p $(j = 1, \ldots, b_p;\, p = 1, \ldots, c)$; otherwise the element is 0. The matrices $\mathbf{Z}_p'\mathbf{Z}_p$ are diagonal.

The modified maximum likelihood method of the present paper can also be applied to this more general class of designs. We again divide the data into two parts with logarithmic likelihoods L' and L'', estimate γ_p and σ^2 by maximizing L', and estimate $\boldsymbol{\alpha}$ by maximizing L''.

With suitable redefinition of $\boldsymbol{\Gamma}$ and $\mathbf{Z}$, results closely follow those already given for the simpler model. Only the main modifications will be described here.

We redefine $\boldsymbol{\Gamma}$ as a diagonal $b \times b$ matrix, i.e. $\boldsymbol{\Gamma} = \operatorname{diag}(\gamma_p\mathbf{I}_p)$ $(p = 1, \ldots, c)$, where $b = \Sigma b_p$ and $\mathbf{I}_p$ is the $b_p \times b_p$ identity matrix; $\mathbf{Z}$ is the partitioned matrix $(\mathbf{Z}_1 \vdots \ldots \vdots \mathbf{Z}_c)$. With these definitions $\mathbf{H}$ can again be written in the form $\mathbf{Z\Gamma Z}'+\mathbf{I}$. The vector $\boldsymbol{\beta}$ and the matrix $\mathbf{U}$ defined by (27) and (28) are also partitioned. The $b_p \times 1$ subvectors of $\boldsymbol{\beta}$ will be denoted by $\boldsymbol{\beta}_p$ and the $b_p \times b_q$ submatrices of $\mathbf{U}$ will be denoted by $\mathbf{U}_{pq}$ $(p, q = 1, \ldots, c)$.

The estimates $\hat{\gamma}_p$ and $\hat{\sigma}^2$ are obtained by solving the c equations

$$-\tfrac{1}{2}E_p+\frac{1}{2\sigma^2}B_p = 0 \quad (p = 1, \ldots, c),$$

together with equation (16), where

$$B_p = \mathbf{y}'(\mathbf{SHS})^{-g}\mathbf{Z}_p\mathbf{Z}_p'(\mathbf{SHS})^{-g}\mathbf{y},$$

$$E_p = \operatorname{tr}\{(\mathbf{SHS})^{-g}\mathbf{Z}_p\mathbf{Z}_p'\mathbf{S}\}.$$

Working expressions for B_p and E_p are

$$B_p = \boldsymbol{\beta}_p' \boldsymbol{\beta}_p / \gamma_p^2, \quad E_p = \operatorname{tr}(\mathbf{U}_{pp}).$$

The information matrix is

$$\tfrac{1}{2}\begin{bmatrix} \operatorname{tr}(\mathbf{U}_{11}^2) & \operatorname{tr}(\mathbf{U}_{12}\mathbf{U}_{21}) & \cdots & \operatorname{tr}(\mathbf{U}_{1c}\mathbf{U}_{c1}) & \operatorname{tr}(\mathbf{U}_{11})/\sigma^2 \\ \operatorname{tr}(\mathbf{U}_{12}\mathbf{U}_{21}) & \operatorname{tr}(\mathbf{U}_{22}^2) & \cdots & \operatorname{tr}(\mathbf{U}_{2c}\mathbf{U}_{c2}) & \operatorname{tr}(\mathbf{U}_{22})/\sigma^2 \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ \operatorname{tr}(\mathbf{U}_{1c}\mathbf{U}_{c1}) & \operatorname{tr}(\mathbf{U}_{2c}\mathbf{U}_{c2}) & \cdots & \operatorname{tr}(\mathbf{U}_{cc})^2 & \operatorname{tr}(\mathbf{U}_{cc})/\sigma^2 \\ \operatorname{tr}(\mathbf{U}_{11})/\sigma^2 & \operatorname{tr}(\mathbf{U}_{22})/\sigma^2 & \cdots & \operatorname{tr}(\mathbf{U}_{cc})/\sigma^2 & (n-t)/\sigma^4 \end{bmatrix}.$$

Generalization of the iterative solution described in § 6 presents little difficulty. Singularities in the information matrix again indicate that one or more parameters cannot be estimated. For example, we cannot estimate γ_q when $\mathbf{U}_{qq} = 0$.

Further generality can be achieved by defining $\mathbf{Z}_p\mathbf{Z}_p'$ as any real symmetric matrix with $\mathbf{Z}_p'\mathbf{Z}_p$ not necessarily diagonal. We can also deal with variance matrices of the form $\mathbf{V} = (\mathbf{J} + \mathbf{Z}\Gamma\mathbf{Z}')\sigma^2$, where $\mathbf{J}$ is symmetric. When $\mathbf{J}$ is idempotent and such that $\mathbf{JZ} = \mathbf{Z}$, $\mathbf{JX} = \mathbf{X}$ and $\mathbf{J} \neq \mathbf{I}$, generalized inverses are required in place of $\mathbf{H}^{-1}$, $(\mathbf{X}'\mathbf{X})^{-1}$ and $(\mathbf{X}'\mathbf{H}^{-1}\mathbf{X})^{-1}$. When $\mathbf{J}$ is not idempotent a preliminary transformation $\mathbf{Ty}$ can be used with $\mathbf{T}$ such that $\mathbf{TJT}'$ is idempotent.

REFERENCES

CUNNINGHAM, E. P. & HENDERSON, C. R. (1968). An iterative procedure for estimating fixed effects and variance components in mixed model situations. *Biometrics* **24**, 13–25.

HARTLEY, H. O. & RAO, J. N. K. (1967). Maximum likelihood estimation for the mixed analysis of variance model. *Biometrika* **54**, 93–108.

HENDERSON, C. R., KEMPTHORNE, O., SEARLE, S. R. & VON KROSIGK, C. M. (1959). The estimation of environmental and genetic trends from records subject to culling. *Biometrics* **15**, 192–218.

NELDER, J. A. (1968). The combination of information in generally balanced designs. *J. R. Statist. Soc.* B **30**, 303–11.

PATTERSON, H. D. (1964). Theory of cyclic rotation experiments (with discussion). *J. R. Statist. Soc.* B **26**, 1–45.

THOMPSON, R. (1969). Iterative estimation of variance components for non-orthogonal data. *Biometrics* **25**, 767–73.

[*Received March* 1971. *Revised June* 1971]

Some key words: Analysis of incomplete block designs; Components of variance; Designs with unequal block sizes; Recovery of inter-block information; Maximum likelihood.

Biometrika (1971), **58**, 3, *p.* 555
Printed in Great Britain

A regression relationship for multivariate paired comparisons

BY ROGER R. DAVIDSON

University of Victoria, British Columbia

AND

RALPH A. BRADLEY

The Florida State University

SUMMARY

In two previous papers, a model for multivariate paired comparisons was proposed and the associated methodology and large-sample properties were developed. The present study considers the problem of relating the response pattern on overall quality to that on a specified set of attributes. A regression equation is derived for a joint distribution of responses to dichotomous items and is applied to the multivariate paired comparison model. A test of the significance of the responses to specified attributes in estimating the responses to overall quality is presented.

1. INTRODUCTION

A multivariate model for paired comparisons has been proposed by Davidson & Bradley (1969). This model is an extension of the univariate model developed by Bradley & Terry (1952) to situations in which there is interest in p attributes or characteristics. The univariate model was proposed first by Zermelo (1929) in discussing chess tournaments. One considers a set of t treatments which are presented in pairs with n_{ij} responses being obtained in the comparison of treatments i and j. Each response to the treatment pair (i,j) consists, in the multivariate case, of a vector of preferences $\mathbf{s} = (s_1, \ldots, s_p)$ whose components s_α indicate which treatment in the pair is preferred on attribute α, that is $s_\alpha = i$ or j $(\alpha = 1, \ldots, p)$. The multivariate model specifies the probability $p(\mathbf{s}|i,j)$ associated with preference vector $\mathbf{s}$ when the treatment pair (i,j) is presented as follows:

$$p(\mathbf{s}|i,j) = p^{(1)}(\mathbf{s}|i,j)\, h(\mathbf{s}|i,j), \tag{1}$$

where

$$p^{(1)}(\mathbf{s}|i,j) = \prod_{\alpha=1}^{p} \pi_{\alpha s_\alpha}/(\pi_{\alpha i} + \pi_{\alpha j}),$$

$$h(\mathbf{s}|i,j) = 1 + \sum_{\alpha<\beta}\sum \delta(s_\alpha, s_\beta)\, \rho_{\alpha\beta} (\pi_{\alpha i}/\pi_{\alpha j})^{\frac{1}{2}\delta(j, s_\alpha)} (\pi_{\beta i}/\pi_{\beta j})^{\frac{1}{2}\delta(j, s_\beta)}, \tag{2}$$

where $s_\alpha = i,j$ $(\alpha = 1, \ldots, p)$ and where $\delta(.,.) = \pm 1$, the sign being positive if the two arguments agree and negative otherwise. The preference parameters,

$$\boldsymbol{\pi} = \{\pi_{\alpha i}\ (i = 1, \ldots, t;\ \alpha = 1, \ldots, p)\},$$

are restrained by

$$\sum_{i=1}^{t} \pi_{\alpha i} = 1 \quad (\alpha = 1, \ldots, p)$$

and the parameters measuring association, $\boldsymbol{\rho} = \{\rho_{\alpha\beta}\ (\alpha < \beta;\ \alpha, \beta = 1, \ldots, p)\}$, are such that