The Annals of Statistics
1976, Vol. 4, No. 1, 86–107

SPANNING SETS FOR ESTIMABLE CONTRASTS IN CLASSIFICATION MODELS[1]

By David Birkes,[2] Yadolah Dodge and Justus Seely

Oregon State University

Two algorithms, the R-process and the Q-process, are presented which can be effective tools for determining the estimable contrasts in a classification model. Both algorithms operate on the incidence matrix of the model as opposed to the design matrix.

If the model is partitioned as $E(Y_u) = h_u \cdot \xi + t_u \cdot \theta$, $u \in U$, the R-process produces a spanning set for the estimable θ-contrasts (i.e., contrasts involving only θ parameters) whenever the set of distinct h_u vectors is linearly independent. If the distinct h_u vectors are dependent, the R-process is still useful and often simplifies the problem of obtaining a spanning set for the estimable θ-contrasts. After the R-process has been applied in a case when the distinct h_u vectors are dependent, the Q-process produces a spanning set for the estimable θ-contrasts provided a partition $h_u \cdot \xi = f_u \cdot \phi + g_u \cdot \omega$, $u \in U$, can be made such that the sets of distinct f_u vectors and distinct g_u vectors are both linearly independent.

As examples, the R-process is used to investigate the additive two-way model; and the R-process and Q-process together are used to investigate an additive three-way model, a two-way model with interaction, and a Graeco-Latin square model.

1. Introduction. For a data set following a classification model it is important to know if the design matrix has maximal rank (or, equivalently, if all cell expectations are estimable). If it has maximal rank then the data set can be analyzed in a straightforward fashion. This is the case when there are observations for all cells or at least for all the cells involved in a known experimental design such as a Latin square design. However, it may happen that an experiment goes awry and the data set does not contain all the observations which were planned. It is in such a situation that the methods of this paper prove useful.

In the event that the design matrix is not of maximal rank, it becomes important to know what parametric functions are estimable when setting up hypotheses to be tested. Of particular interest is finding what contrasts involving only a single effect are estimable. It is the hypothesis that these estimable contrasts are zero which is tested by the F-ratio formed from the sum of squares

Received October 1974; revised May 1975.

[1] This paper is a combination of two papers, the first by Seely, Birkes, Dodge, and Hartmann and the second by Birkes, Dodge, and Seely.

[2] Research supported by NSF grant GJ-499 and NIH grant ES 00210.

AMS 1970 *subject classifications.* Primary 62J99; Secondary 62K99.

Key words and phrases. Classification models, missing observations, estimable contrasts, degrees of freedom, maximal rank design matrices.

for that effect adjusted for all other effects. Furthermore, the number of linearly independent such estimable contrasts is the appropriate numerator degrees of freedom associated with the F-ratio. Also, a knowledge of the estimable parametric functions allows one to easily obtain full rank reparametrizations needed for some computer programs.

In dealing with a general linear model, one usually investigates estimability directly in terms of the design matrix. For the case of a classification model the same information contained in the design matrix can also be derived from the incidence matrix. We take this latter approach.

Utilizing the incidence matrix to obtain estimability information seems to have been first considered by Bose [3] for the additive two-way model. After introducing a notion of connectedness based on the incidence matrix, Bose proved that a design is connected if and only if all contrasts within each factor are estimable, which is equivalent to the design matrix having maximal rank. One would like to be able to formulate a definition for the connectedness of a design for other classification models such that (1) the connectedness of a design is easily determined from the incidence matrix and (2) the theorem is true that a design is connected if and only if the associated design matrix has maximal rank. Weeks and Williams [11] define a relationship between design points (which correspond to occupied cells in the incidence matrix) in an additive multi-way model which leads to a sufficient condition for a design matrix to have maximal rank; but as noted in [11] the condition is not necessary. Srivastava and Anderson [10] also deal with additive multi-way models; they obtain a necessary and sufficient condition on the incidence matrix for a design matrix to be of maximal rank. However, it is not clear that this condition could be verified by an algorithm which would terminate after a specified number of steps. We know of no completely satisfactory generalization of Bose's result even to the additive three-way model.

2. The problem. Let $\{Y_u : u \in U\}$ be a collection of random variables where U is a finite nonempty index set. We assume these random variables have expectations of the form

$$E(Y_u) = h_u \cdot \xi + t_u \cdot \theta \,, \quad u \in U \,,$$

where ξ and θ are column vectors of parameters and h_u and t_u are column vectors of known real numbers. (Although a prime will be used to denote the transpose of a vector or matrix, here we use the dot product notation $h_u \cdot \xi$ rather than $h_u'\xi$.) The parameters are assumed to be unrestricted; however, as indicated in Section 9 our results can be useful even when there are known linear constraints on the parameter vectors.

We use the terms "linear parametric function" and "estimable" in the standard fashion. That is, a *linear parametric function* is a linear combination of the parameters in ξ and θ, or more precisely, a linear functional on the vector space of parameter vectors $(\xi', \theta')'$; and a linear parametric function is said to be *estimable* provided it can be written as a linear combination of the expectations $h_u \cdot \xi + t_u \cdot \theta, u \in U$. A linear parametric function involving only the parameters in θ will be called a *θ-functional*; and when the sum of its coefficients is zero it will be called a *θ-contrast*.

In this general setting the problem is to find a procedure which will yield a spanning set for the vector space of estimable θ-functionals. Although some of our results are obtained for general partitioned linear models, our underlying concern throughout is with classification models.

In an n-way classification model the expectation of each random variable is determined by an n-tuple $(i_1, \cdots, i_n)$ where the jth component indicates that the random variable is associated with the i_jth level of the jth factor. The parameters involved in these expectations are the grand mean, n groups of main effects and, depending on the particular model, certain groups of interaction effect and nested effect parameters. The *incidence matrix* of the model is the n-dimensional matrix whose entry in the $(i_1, \cdots, i_n)$ position, or *cell*, is the number of random variables in our collection associated with that n-tuple. If all the cells of the incidence matrix are occupied (i.e., have nonzero entries), then of course all the usual linear parametric functions are estimable. When there are "missing observations" (i.e., unoccupied cells), we would like to be able to determine from the pattern of the occupied cells which of the usual linear parametric functions are still estimable.

This paper presents two algorithms, the R-process and the Q-process, which appear to be useful tools for working on estimability problems in classification models. These two algorithms can be applied to any classification model, but are most effective whenever the hypotheses of Proposition 4.1 or Theorem 7.4 are satisfied.

3. The R-process. The *R-process* is a procedure applied to a two-dimensional matrix W with nonnegative integers as entries to obtain a matrix M of the same size with zeros and ones as entries. It is convenient to allow the rows and columns of W to be indexed by any two finite index sets (i.e., not necessarily sets of integers), say R and C, and to denote the entries of W by w_{rc} for $r \in R$ and $c \in C$. (The rows, columns, and entries of M are to be indexed exactly like W.) The R-process applied to the matrix W is defined as follows:

1) For all $r \in R$ and $c \in C$, set m_{rc} equal to zero or one according as w_{rc} is zero or nonzero.

2) Change any zero m_{rc} to one if there exists $s \in R$ and $d \in C$ such that $m_{rd} = m_{sd} = m_{sc} = 1$. (Pictorially, we add the fourth corner whenever three corners of a rectangle appear in the matrix.)

3) Continue step 2, using both the original and the new nonzero m_{rc}'s as corners of new rectangles, until no more entries can be changed.

DEFINITION 3.1. The matrix M which results from applying the R-process to a two-dimensional matrix W is called the *final matrix* obtained from W.

An interesting alternative description of the R-process is the following: Compute the sequence of matrices

$$W_0 = W, \qquad W_1 = W_0 W_0' W_0, \qquad W_2 = W_1 W_1' W_1,$$

and so on, until it happens that two consecutive matrices W_l and W_{l+1} have the same number of nonzero entries. Then form M by letting m_{rc} equal zero or one according as the (r, c) entry of W_l is zero or nonzero. To see that this matrix M is the same as the final matrix obtained previously, let $w_{rc}^{(i)}$ denote the (r, c) entry of W_i in the above sequence and observe that

$$w_{rc}^{(i+1)} = \sum_{s \in R} \sum_{d \in C} w_{rd}^{(i)} w_{sd}^{(i)} w_{sc}^{(i)} .$$

We mention without proof that l satisfies the inequality $(3^{l-1} + 1)/2 < \min\{a, b\}$ where a and b are the numbers of rows and columns of W respectively.

ASSUMPTION. Hereafter we will make the assumption that each row and column of W has at least one nonzero entry. This simplifies the definition of the equivalence relation $\sim$ which follows. Notice that the matrix W which is defined later in this section, when we treat the model introduced in Section 2, satisfies the assumption.

DEFINITION 3.2. In the set C indexing the columns of W and M we say c is *equivalent* to d and write $c \sim d$ if the two columns of M indexed by c and d are identical.

It is clear that the relation $\sim$ is an equivalence relation on C. Let $C_1, \cdots, C_q$ be the distinct equivalence classes in C. By interchanging the roles of rows and columns we could define an equivalence relation on R. It is not hard to verify that the equivalence classes in R can be described as the q sets $R_k = \{r \in R : m_{rc} = 1$ for some $c \in C_k\}$ for $k = 1, \cdots, q$. Therefore, $m_{rc} = 1$ if and only if there is some k, $1 \leqq k \leqq q$, such that $r \in R_k$ and $c \in C_k$.

From the first description of the R-process it can be seen that $c \sim d$ if and only if $m_{rc} = m_{rd} = 1$ for some $r \in R$. Consider the matrix $M'M$. Its rows and columns are both indexed by C, and its (c, d) entry is $\sum_{r \in R} m_{rc} m_{rd}$. Therefore:

LEMMA 3.3. *$c \sim d$ if and only if the (c, d) entry of $M'M$ is nonzero.*

REMARK 3.4. Another way to arrive at this equivalence relation can be found in Bose [3]. One would employ a search procedure based on the original matrix W, as illustrated on page 324 of Searle [8]. It seems, however, that our procedure based on the final matrix obtained by the R-process is faster and easier to implement.

We now consider some implications of the R-process with respect to the partitioned linear model in Section 2. Let $\mathscr{E}$ denote the set of all estimable linear parametric functions for our model. From the definition of "estimable" it is clear that $\mathscr{E}$ is a vector space and that $\mathscr{E}$ is spanned by $\{h_u \cdot \xi + t_u \cdot \theta : u \in U\}$. Let H and T be the sets of distinct vectors among the h_u's and t_u's respectively. Define a two-dimensional matrix W, with rows indexed by H and columns indexed by T, by letting the entry w_{ht} be the number of indices $u \in U$ such that $(h_u, t_u) = (h, t)$. (For an n-way classification model the matrix W is easily formed from the cells of the associated n-dimensional incidence matrix.) Let M be the final matrix obtained from W by the R-process.

PROPOSITION 3.5. *If* $m_{ht} = 1$, *then* $h \cdot \xi + t \cdot \theta$ *is estimable.*

PROOF. If $m_{ht} = 1$ at the beginning of the R-process, then for some $u \in U$, $h \cdot \xi + t \cdot \theta = h_u \cdot \xi + t_u \cdot \theta$, which of course is estimable. As the R-process proceeds, if m_{ht} is set equal to 1 it is because there are $g \in H$ and $s \in T$ such that $m_{hs} = m_{gs} = m_{gt} = 1$. Now observe that

$$h \cdot \xi + t \cdot \theta = (h \cdot \xi + s \cdot \theta) - (g \cdot \xi + s \cdot \theta) + (g \cdot \xi + t \cdot \theta) . \qquad \square$$

An immediate consequence of Proposition 3.5 is that the set $E = \{h \cdot \xi + t \cdot \theta : m_{ht} = 1\}$ is a spanning set for $\mathscr{E}$. It is generally handier to work with E rather than the spanning set $\{h_u \cdot \xi + t_u \cdot \theta : u \in U\}$. Indeed, for estimability considerations we may as well assume we have a model with expectations forming the set E.

PROPOSITION 3.6. *If* $t_1 \sim t_2$, *then* $(t_1 - t_2) \cdot \theta$ *is estimable.*

PROOF. There is some $h \in H$ such that $m_{ht_1} = m_{ht_2} = 1$. By Proposition 3.5, $h \cdot \xi + t_1 \cdot \theta$ and $h \cdot \xi + t_2 \cdot \theta$ are estimable; and hence so is their difference. $\square$

Choose a complete set S of representatives for the equivalence classes in T. (S is formed by choosing one element from each of the equivalence classes.) Two facts we will be using later are that for each $h \in H$ there is exactly one $s \in S$ with $m_{hs} = 1$ and that $m_{hs} = m_{ht}$ for all $t \sim s$. Let

$$D = \{(t - s) \cdot \theta : t \in T, s \in S, t \sim s, t \neq s\} ,$$
$$E^* = \{h \cdot \xi + s \cdot \theta : h \in H, s \in S, m_{hs} = 1\} ;$$

and let $\mathscr{D}$ and $\mathscr{E}^*$ denote the vector spaces spanned by D and E^* respectively.

PROPOSITION 3.7. $\mathscr{E} = \mathscr{D} + \mathscr{E}^*$.

PROOF. Propositions 3.5 and 3.6 imply that $\mathscr{D} + \mathscr{E}^* \subset \mathscr{E}$. Now let $e \in \mathscr{E}$. Since E spans $\mathscr{E}$ we can write $e = \sum_{h \in H} \sum_{t \in T} c_{ht}(h \cdot \xi + t \cdot \theta)$ where $c_{ht} = 0$ if $m_{ht} = 0$. Now rewrite e as

$$\begin{aligned} e &= \textstyle\sum_{h \in H} \sum_{s \in S} \sum_{t \sim s} c_{ht}(h \cdot \xi + t \cdot \theta) \\ &= \textstyle\sum_{h \in H} \sum_{s \in S} \sum_{t \sim s} [c_{ht}(h \cdot \xi + s \cdot \theta) + c_{ht}(t - s) \cdot \theta] \\ &= \textstyle\sum_{h \in H} \sum_{s \in S} (\sum_{t \sim s} c_{ht})(h \cdot \xi + s \cdot \theta) + \sum_{h \in H} \sum_{s \in S} \sum_{t \sim s} c_{ht}(t - s) \cdot \theta . \end{aligned}$$

Clearly the second term is in $\mathscr{D}$. If $m_{hs} = 0$, then $\sum_{t \sim s} c_{ht} = 0$ since $m_{ht} = m_{hs}$ for all $t \sim s$. This shows that the first term is in $\mathscr{E}^*$. $\square$

Let $\mathscr{E}_\theta$ denote the set of estimable θ-functionals for our model, i.e., the set of θ-functionals in $\mathscr{E}$, and let $\mathscr{E}_\theta{}^*$ denote the set of θ-functionals in $\mathscr{E}^*$. Clearly both $\mathscr{E}_\theta$ and $\mathscr{E}_\theta{}^*$ are vector spaces. Moreover:

PROPOSITION 3.8. $\mathscr{E}_\theta = \mathscr{D} + \mathscr{E}_\theta{}^*$.

PROOF. This follows from the preceding proposition and the fact that $\mathscr{D} \subset \mathscr{E}_\theta$. □

With respect to finding a spanning set for $\mathscr{E}_\theta$, the propositions above indicate that we can first use the R-process to obtain the set D and then restrict attention to finding a spanning set for $\mathscr{E}_\theta{}^*$. In this connection it is sometimes useful to regard $\mathscr{E}_\theta{}^*$ as the vector space of estimable θ-functionals from a "reduced model" with expectations forming the set E^*.

REMARK 3.9. It may happen that, for a reason other than Proposition 3.6, we know some θ-functional $(t_1 - t_2) \cdot \theta$ is estimable. For example, perhaps $(t_1 - t_2) \cdot \theta = (v_1 - v_2) \cdot \theta$ where $v_1 \sim v_2$. We could then modify the R-process by inserting a step between steps 1 and 2 in which we would set $m_{ht_1} = 1$ whenever $m_{ht_2} = 1$, and vice versa. All of the results of this section would remain true. Such a modification will sometimes decrease (and never increase) the size of S, which will usually reduce the effort required to find a spanning set for $\mathscr{E}_\theta$.

REMARK 3.10. Once a spanning set for $\mathscr{E}_\theta$ is obtained, one will often proceed to extract a basis for $\mathscr{E}_\theta$ (see Statement (9.8)). In this regard it is helpful to know that when the vectors in T are linearly independent, then $\mathscr{D} \cap \mathscr{E}_\theta{}^* = \{0\}$ (in fact, $\mathscr{D} \cap \mathscr{E}^* = \{0\}$) and D is a basis for $\mathscr{D}$. (See the proof of Proposition 4.3.)

4. Additive two-way classification model. We have seen how the R-process can be used as a first step in finding a spanning set for $\mathscr{E}_\theta$. For an additive two-way classification model we will see that the R-process suffices to solve the entire problem. More generally:

PROPOSITION 4.1. *Suppose the vectors in H are linearly independent. Then $\mathscr{E}_\theta$ is spanned by D*, i.e., $\mathscr{E}_\theta = \mathscr{D}$.

PROOF. Suppose $e \in \mathscr{E}_\theta$. Because E spans $\mathscr{E}$ and because ξ and θ are unrestricted, we can write $e = \sum_h \sum_t c_{ht} t \cdot \theta$ where the c_{ht}'s are such that $\sum_h \sum_t c_{ht} h = 0$ and $c_{ht} = 0$ whenever $m_{ht} = 0$. The linear independence implies $\sum_t c_{ht} = 0$ for all $h \in H$. For $h \in H$ let $s(h)$ denote the unique element $s \in S$ such that $m_{hs} = 1$; and note that c_{ht} can be nonzero only when $t \sim s(h)$. As a consequence

$$e = \textstyle\sum_{h \in H} \sum_{t \sim s(h)} c_{ht} t \cdot \theta$$

and

$$\textstyle\sum_{t \sim s(h)} c_{ht} = \sum_t c_{ht} = 0 \qquad \text{for all} \quad h \in H\,.$$

Hence

$$\begin{aligned} e &= \textstyle\sum_{h \in H} [\sum_{t \sim s(h)} c_{ht} t \cdot \theta - (\sum_{t \sim s(h)} c_{ht}) s(h) \cdot \theta] \\ &= \textstyle\sum_{h \in H} \sum_{t \sim s(h)} c_{ht}(t - s(h)) \cdot \theta\,. \end{aligned}$$

Since the terms involving $t = s(h)$ in the above sum can be disregarded, it follows that $e \in \mathscr{D}$. □

Now consider the additive two-way classification model

$$E(Y_{ijk}) = \mu + \alpha_i + \beta_j ,$$

where $i = 1, \cdots, a$, $j = 1, \cdots, b$, and $k = 1, \cdots, n_{ij}$, with the usual interpretation that when $n_{ij} = 0$ there are no random variables with the first two subscripts i, j. The incidence matrix is $N = (n_{ij})$, which we suppose has no row nor column composed completely of zeros. Suppose we are interested in finding all estimable β-functionals. To put the model in the form we have been dealing with, let $U = \{(i, j, k) : i = 1, \cdots, a; j = 1, \cdots, b; k = 1, \cdots, n_{ij}\}$, $\xi = (\mu, \alpha_1, \cdots, \alpha_a)'$, $\theta = (\beta_1, \cdots, \beta_b)'$, and for $u = (i, j, k) \in U$ let h_u be an $(a + 1) \times 1$ vector with its first and $(i + 1)$th components 1 and all other components zero, and let t_u be a $b \times 1$ vector with its jth component 1 and all others zero.

For indexing the rows and columns of N we use the sets $I = \{1, \cdots, a\}$ and $J = \{1, \cdots, b\}$ respectively; that is, we use i in place of h_u and j in place of t_u where $u = (i, j, k)$. Apply the R-process to N to obtain the final matrix M. From Definition 3.2 and the paragraph following it, we have equivalence relations on the sets I and J; let the equivalence classes be $I_1, \cdots, I_r$ and $J_1, \cdots, J_r$ respectively, where the subscripting is such that $m_{ij} = 1$ if and only if $i \in I_k$ and $j \in J_k$ for some k, $1 \leqq k \leqq r$. (In the terminology of Bose [3], the pairs $(I_1, J_1), \cdots, (I_r, J_r)$ describe the r connected portions of the design.) From Proposition 3.5 we know that each element in $E = \{\mu + \alpha_i + \beta_j : m_{ij} = 1\}$ is estimable. Moreover:

PROPOSITION 4.2. *The cell expectation* $\mu + \alpha_i + \beta_j$ *is estimable if and only if* $m_{ij} = 1$.

PROOF. When $m_{ij} = 1$, apply Proposition 3.5. Conversely, suppose $\mu + \alpha_i + \beta_j$ is estimable. Then we can write $\mu + \alpha_i + \beta_j = \sum_p \sum_q c_{pq}(\mu + \alpha_p + \beta_q)$ where $c_{pq} = 0$ if $m_{pq} = 0$. Let k be such that $i \in I_k$. We must show that $j \in J_k$. From the linear independence of the α_p's we see $\sum_{q=1}^{b} c_{pq}$ equals 1 if $p = i$ and equals 0 if $p \neq i$. Note that $\sum_{q=1}^{b} c_{pq} = \sum_{q \in J_k} c_{pq}$ for all $p \in I_k$ and $\sum_{p=1}^{a} c_{pq} = \sum_{p \in I_k} c_{pq}$ for all $q \in J_k$. Now

$$\begin{aligned} 1 &= \textstyle\sum_{p \in I_k} (\sum_{q=1}^{b} c_{pq}) = \sum_{p \in I_k} (\sum_{q \in J_k} c_{pq}) \\ &= \textstyle\sum_{q \in J_k} (\sum_{p \in I_k} c_{pq}) = \sum_{q \in J_k} (\sum_{p=1}^{a} c_{pq}) . \end{aligned}$$

From the linear independence of the β_q's we see $\sum_{p=1}^{a} c_{pq} = 0$ for $q \neq j$. Therefore, we must have $j \in J_k$. □

The set D for the additive two-way model is

$$D = \{\beta_j - \beta_{j_k} : j \in J_k, j \neq j_k, k = 1, \cdots, r\} ,$$

where j_k is a fixed element of J_k. It is clear that Proposition 4.1 is applicable and hence D spans $\mathscr{E}_\beta$, the vector space of estimable β-contrasts. (Note that an estimable β-functional is necessarily a β-contrast.) In fact:

PROPOSITION 4.3. *For the additive two-way model, the set D is a basis for $\mathscr{E}_\beta$.*

PROOF. It only remains to show that the elements in D are linearly independent. But this follows because the β_j's are all linearly independent and because the β_j terms in the elements $\beta_j - \beta_{j_k}$ of D are all distinct. □

The dimension d_β of $\mathscr{E}_\beta$ is generally referred to as the degrees of freedom associated with the sum of squares for the β-effects adjusted for μ and α-effects. Counting the elements in D we see that $d_\beta = b - r$. Therefore, the dimension d of $\mathscr{E}$ is equal to $a + d_\beta = a + b - r$. (See Section 9 below or Section 7.4 in [8].)

PROPOSITION 4.4. *A β-contrast $\beta_k - \beta_l$ is estimable if and only if the (k, l) entry of $M'M$ is nonzero.*

PROOF. The "if" part follows from Lemma 3.3 and Proposition 3.6. Conversely, suppose $\beta_k - \beta_l$ is estimable. Since the kth column of N is nonzero, $m_{pk} = 1$ for some p. But then $\mu + \alpha_p + \beta_l = (\mu + \alpha_p + \beta_k) - (\beta_k - \beta_l)$ is estimable, and so $m_{pl} = 1$ by Proposition 4.2. Hence $m_{pk}m_{pl} = 1$, so $\sum_i m_{ik}m_{il} \neq 0$. □

One method for obtaining a basis for the estimable β-contrasts makes use of the *contrast triangle* Δ, which is the triangular portion of $M'M$ below the diagonal, with all nonzero entries replaced by ones. That is,

$$\begin{aligned} \Delta_{kl} &= 1 \quad \text{if} \quad \textstyle\sum_i m_{ik}m_{il} \neq 0\,, \\ &= 0 \quad \text{if} \quad \textstyle\sum_i m_{ik}m_{il} = 0\,, \qquad 1 \leqq l < k \leqq b\,. \end{aligned}$$

Now d_β is simply the number of nonzero rows in Δ. And a basis for $\mathscr{E}_\beta$ can be formed by selecting one contrast $\beta_k - \beta_l$ such that $\Delta_{kl} = 1$ for each nonzero row k. This basis is generally different from the basis D.

Let us justify the above method. That the contrasts selected are estimable follows from Proposition 4.4. The kth row of Δ is zero if and only if there is no $l < k$ such that $k \sim l$. This is true precisely for the smallest integers in the sets J_q, $1 \leqq q \leqq r$. Therefore there are $d_\beta = b - r$ nonzero rows in Δ. It remains to show that the contrasts $\beta_{k'} - \beta_l$ selected above are linearly independent. Given a linear combination of these contrasts with at least one nonzero coefficient, consider the largest index k such that the coefficient of $\beta_k - \beta_l$, say c_{kl}, is nonzero. When this linear combination is expressed as a sum of β_j terms, the coefficient of β_k is $c_{kl} \neq 0$. Thus the sum cannot be zero, because the β_j's are linearly independent.

All the above results on β-contrasts can be easily translated into results about α-contrasts. For instance, a basis for the estimable α-contrasts can be obtained by using the triangular portion of MM' below the diagonal.

As an example let us consider an additive two-way model with $a = 5$, $b = 6$, and an incidence matrix $N = (n_{ij})$ whose nonzero entries occur in the cells occupied by 1's in the following matrix:

	1	2	3	4	5	6
1	1	x			1	
2			x	1		x
3	x	1			1	
4			1	x		1
5			x	1		1

The x's are in those cells which become occupied after application of the R-process. That is, the final matrix M is the 5×6 matrix with 1's in those cells above which are occupied by 1's and x's. The contrast triangles for α-contrasts and for β-contrasts are

$\Delta_\alpha =$

	1	2	3	4
2	0			
3	1	0		
4	0	1	0	
5	0	1	0	1

$\Delta_\beta =$

	1	2	3	4	5
2	1				
3	0	0			
4	0	0	1		
5	1	1	0	0	
6	0	0	1	1	0

From Δ_β we seen that $d_\beta = 4$, $d = a + d_\beta = 9$, and that a basis for the estimable β-contrasts is $\{\beta_1 - \beta_2, \beta_3 - \beta_4, \beta_1 - \beta_5, \beta_3 - \beta_6\}$. This already tells us $d_\alpha = d - b = 3$. To get a basis for the estimable α-contrasts we must look at Δ_α, which shows us $\{\alpha_1 - \alpha_3, \alpha_2 - \alpha_4, \alpha_2 - \alpha_5\}$ is such a basis.

REMARK. Results similar to some of the facts in this section have recently been given by Eccleston and Hedayat (see Section 3A in [5]).

5. Further applications of the R-process. An n-way classification model with $n > 2$ allows several possible "natural" partitions of the parameters into two subvectors. For any one of these partitions application of the R-process may tell us that the cell expectations associated with some of the unoccupied cells of the incidence matrix are estimable (see Proposition 3.5). As far as estimability is concerned, we can suppose that these cells are occupied. Now we can use this new incidence matrix and a different partition of the parameters to try to discover more estimable cell expectations. If this leads to the discovery that all cell expectations are estimable, then the R-process by itself will have answered all our questions about estimability.

The method presented by Weeks and Williams [11] (or see page 338 in [8]) for dealing with an additive n-way classification model can be reformulated in terms of the R-process. For each of the n factors partition the parameters into those which represent the effects of the levels of that factor and those which do not. Use these partitions, in any fixed order, as indicated in the preceding paragraph; that is, after each partition is used, new occupied cells are added to the incidence matrix if their cell expectations are found to be estimable. Continue to pass through these n steps until n consecutive steps fail to change the incidence matrix. Suppose α_1 and α_2 are parameters representing two levels of one factor. If the two $(n - 1)$-dimensional matrices of cells in the current incidence matrix whose expectations involve α_1 and α_2, respectively, have occupied cells in at least one common position (or equivalently, if the two matrices are identical and nonzero), then $\alpha_1 - \alpha_2$ is estimable. Such estimable contrasts are precisely the ones found in [11]. The actual method described in [11] employs a search procedure based on the original matrix, and a comment similar to Remark 3.4 applies.

Using partitions other than the n partitions mentioned above can lead to a more effective procedure as is seen in the following example. Consider the additive four-way classification model given by

$$E(Y_{ijkpq}) = \mu + \alpha_i + \beta_j + \gamma_k + \delta_p ,$$

where $i, p = 1, 2, j, k = 1, 2, 3$, and $q = 1, \cdots, n_{ijkp}$. Suppose the incidence matrix $N = (n_{ijkp})$ is the $2 \times 3 \times 3 \times 2$ matrix represented by the following configuration of cells:

$$\tag{5.1} \begin{array}{ccc} & & \begin{array}{ccc} & \gamma_1 & \\ \beta_1 & \beta_2 & \beta_3 \end{array} \quad \gamma_2 \quad\quad \gamma_3 \\ \delta_1 & \begin{array}{c}\alpha_1\\ \alpha_2\end{array} & \begin{bmatrix}1&0&0\\0&1&0\end{bmatrix}\begin{bmatrix}0&0&1\\0&0&0\end{bmatrix}\begin{bmatrix}0&1&0\\0&0&0\end{bmatrix} \\ \delta_2 & & \begin{bmatrix}0&0&0\\0&0&0\end{bmatrix}\begin{bmatrix}1&0&0\\0&0&0\end{bmatrix}\begin{bmatrix}0&0&1\\0&1&0\end{bmatrix}. \end{array}$$

Thus, for example, $n_{1321} = 1$ and $n_{2212} = 0$.

If we use only the four partitions corresponding to single factors, we are not able to discover that the design matrix associated with this incidence matrix has maximal rank. However, let us partition the parameters into the subvectors $(\mu, \alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3)'$ and $(\gamma_1, \gamma_2, \gamma_3, \delta_1, \delta_2)'$. Now an application of the R-process tells us that we can assume the pattern of occupied cells to be:

$$\tag{5.2} \begin{array}{c} \begin{bmatrix}1&0&1\\0&1&0\end{bmatrix}\begin{bmatrix}1&0&1\\0&1&0\end{bmatrix}\begin{bmatrix}0&1&0\\0&0&0\end{bmatrix} \\ \begin{bmatrix}0&0&0\\0&0&0\end{bmatrix}\begin{bmatrix}1&0&1\\0&1&0\end{bmatrix}\begin{bmatrix}1&0&1\\0&1&0\end{bmatrix}. \end{array}$$

At this point we can use the partition corresponding to the δ factor followed by the partition corresponding to the α factor to find that all cell expectations are estimable. Alternatively, this fact could be obtained from (5.2) by using the partition corresponding to the α and δ factors versus the β and γ factors.

REMARK 5.3. If the incidence matrix is represented by a configuration such as in the above example, then for certain partitions of the parameters the new incidence matrix obtained by applying the R-process may be obtained more quickly by directly using the configuration. To illustrate, consider the incidence matrix N represented by (5.1) and the partition leading to (5.2). One way of forming W, the matrix to which the R-process is applied, is to form a 6×6 matrix whose columns are formed from the six 2×3 submatrices of (5.1) and then to eliminate all zero rows and columns. The R-process can be performed on W in the following manner: Find pairs of columns of W which have nonzero entries in the same row and then make both columns the same by setting an entry of either column equal to 1 whenever the corresponding entry in the other column is nonzero, and continue this procedure until no more entries of W can be changed. From these descriptions of the formation of W and the application of the R-process it is clear that we can write down (5.2) directly from (5.1) by comparing corresponding cells in the six 2×3 submatrices. For example, since $n_{2211} = n_{2232} = 1$, we can set n_{1311} and n_{1132} equal to 1.

6. Loops. In Section 3 the problem of finding a spanning set for the estimable θ-functionals for the model $E(Y_u) = h_u \cdot \xi + t_u \cdot \theta$, $u \in U$, was reduced to the problem of finding a spanning set for $\mathscr{E}_\theta{}^*$, the θ-functionals in the vector space $\mathscr{E}^*$ spanned by E^*. Progress toward a solution of the latter problem often can be made by partitioning ξ into two subvectors ϕ and ω. Then we can write

$$E(Y_u) = f_u \cdot \phi + g_u \cdot \omega + t_u \cdot \theta\,, \quad u \in U\,,$$

where $\xi = (\phi', \omega')'$ and $h_u = (f_u', g_u')'$.

Let F and G denote the sets of distinct f_u's and g_u's respectively. For each $s \in S$ select some symbol Ω_s which is not the numeral 0. Define a two-dimensional matrix $\bar{M}$, with rows indexed by F and columns indexed by G, by defining

$$\begin{aligned} \bar{m}_{fg} &= \Omega_s \qquad \text{if } \; f \cdot \phi + g \cdot \omega + s \cdot \theta \in E^* \quad \text{for some} \quad s \in S \\ &= 0 \qquad \text{otherwise.} \end{aligned}$$

The matrix $\bar{M}$ is well-defined because $f \cdot \phi + g \cdot \omega + s \cdot \theta \in E^*$ if and only if $(f', g')' = h \in H$ and $m_{hs} = 1$ and because for each $h \in H$ there is only one $s \in S$ such that $m_{hs} = 1$. The use of the symbols Ω_s is a device which can help us remember which vector s is associated with which pair (f, g).

Another way to view the formation of $\bar{M}$ is as follows: For each $s \in S$ rearrange the entries of the column of M indexed by s into a two-dimensional matrix $M^{(s)}$ with rows indexed by F and columns indexed by G. Specifically, let $m_{fg}^{(s)} = m_{hs}$ if $(f', g')' = h \in H$ and let $m_{fg}^{(s)} = 0$ if $(f', g')' \notin H$. Replace each entry 1 in $M^{(s)}$ by Ω_s for all $s \in S$ and then "collapse" these matrices together into one matrix $\bar{M}$.

DEFINITION 6.1. A *loop* in the matrix $\bar{M}$ is a sequence of an even number n ($n > 0$) of pairs $(f_1, g_1)(f_2, g_2) \cdots (f_n, g_n)$ such that

(i) the pairs are distinct,
(ii) the (f_i, g_i) entry of $\bar{M}$ is nonzero for all i,

(iii) $f_i = f_{i+1}$ for i odd ($i = 1, 3, \cdots, n-1$),
(iv) $g_i = g_{i+1}$ for i even ($i = 2, 4, \cdots, n-2$),
(v) $g_1 = g_n$.

(If we drew lines between the entries corresponding to successive pairs in a loop, we would obtain a picture of a rectilinear loop.)

For a loop L such as in Definition 6.1 let $\theta(L)$ denote the θ-functional

$$\theta(L) = (s_1 - s_2 + \cdots - s_n) \cdot \theta$$

where s_i is the unique element of S such that $f_i \cdot \phi + g_i \cdot \omega + s_i \cdot \theta \in E^*$ (thus Ω_{s_i} is the (f_i, g_i) entry of $\bar{M}$). Observe that $\theta(L) = \sum_{i=1}^{n} (-1)^{i+1}(f_i \cdot \phi + g_i \cdot \omega + s_i \cdot \theta)$. Therefore:

PROPOSITION 6.2. $\theta(L) \in \mathscr{E}_\theta^*$.

Let Λ be the set of all loops in $\bar{M}$. We just saw that $\theta(L) \in \mathscr{E}_\theta^*$ for all $L \in \Lambda$. These θ-functionals will span $\mathscr{E}_\theta^*$ in certain cases, such as when $\phi = (\mu, \alpha_1, \cdots, \alpha_a)'$ and $\omega = (\beta_1, \cdots, \beta_b)'$ where α and β represent two main effects in a classification model. More generally:

LEMMA 6.3. *Suppose the vectors in F are linearly independent and the vectors in G are linearly independent. Then $\mathscr{E}_\theta^*$ is spanned by $\{\theta(L) : L \in \Lambda\}$.*

PROOF. Suppose $a \cdot \theta$ is a nonzero element of $\mathscr{E}_\theta^*$. Then $a \cdot \theta = \sum_{h \in H} \sum_{s \in S} c_{hs}(h \cdot \xi + s \cdot \theta)$ where $c_{hs} = 0$ if $m_{hs} = 0$. For $f \in F$, $g \in G$, $s \in S$ let us set $d_{fgs} = c_{hs}$ if $(f', g')' = h \in H$ and $d_{fgs} = 0$ otherwise. Now we can write

$$a \cdot \theta = \sum_{f \in F} \sum_{g \in G} \sum_{s \in S} d_{fgs}(f \cdot \phi + g \cdot \omega + s \cdot \theta),$$

where $d_{fgs} = 0$ if $\bar{m}_{fg} = 0$, i.e., if $f \cdot \phi + g \cdot \omega + s \cdot \theta \notin E^*$. Set $D_{fg} = \sum_{s \in S} d_{fgs}$ and note for each pair (f, g) there is at most one $s \in S$ such that $d_{fgs} \neq 0$. By the hypotheses of the lemma and the fact that ϕ, ω and θ are unrestricted, the linear parametric functions $f \cdot \phi$ for $f \in F$ and $g \cdot \omega$ for $g \in G$ are linearly independent of one another and of all θ-functionals. Therefore, $\sum_{g \in G} D_{fg} = 0$ for all $f \in F$ and $\sum_{f \in F} D_{fg} = 0$ for all $g \in G$.

There must be some nonzero coefficient $d_{f_1 g_1 s_1}$; then $D_{f_1 g_1} = d_{f_1 g_1 s_1} \neq 0$. Since $\sum_g D_{f_1 g} = 0$, there is some $g_2 \neq g_1$ with $D_{f_1 g_2} \neq 0$. Since $\sum_f D_{f g_2} = 0$, there is some $f_2 \neq f_1$ with $D_{f_2 g_2} \neq 0$. In this way we get a sequence $(f_1, g_1)(f_1, g_2)(f_2, g_2)(f_2, g_3)(f_3, g_3) \cdots$, where $f_p \neq f_{p+1}$ and $g_p \neq g_{p+1}$ for all p and $D_{fg} \neq 0$, hence $\bar{m}_{fg} \neq 0$, for each pair (f, g) in the sequence. Let r be the first integer such that $g_q = g_r$ for some $q < r$. If $f_{r-1} \neq f_q$, then $(f_q, g_q)(f_q, g_{q+1})(f_{q+1}, g_{q+1}) \cdots (f_{r-1}, g_{r-1})(f_{r-1}, g_r)$ is a loop in $\bar{M}$. If $f_{r-1} = f_q$, then $(f_{r-1}, g_{r-1})(f_q, g_{q+1})(f_{q+1}, g_{q+1}) \cdots (f_{r-2}, g_{r-2})(f_{r-2}, g_{r-1})$ is a loop in $\bar{M}$. In either case let us, for convenience, re-index the pairs and write the loop as $L = (f_1, g_1)(f_2, g_2) \cdots (f_n, g_n)$, where $n = 2r - 2q$ in the first case and $n = 2r - 2q - 2$ in the second case.

For each (f_i, g_i) in L let s_i be the unique element of S such that $D_{f_i g_i} = d_{f_i g_i s_i} \neq 0$. We observed previously that $\theta(L) = \sum_{i=1}^{n} (-1)^{i+1}(f_i \cdot \phi + g_i \cdot \omega + s_i \cdot \theta)$. Now $a \cdot \theta - d_{f_1 g_1 s_1}\theta(L) \in \mathscr{E}_\theta^*$ and it can be expressed as a linear combination of elements of E^* with strictly fewer nonzero coefficients than the expression for $a \cdot \theta$. Using induction we can argue that every element of $\mathscr{E}_\theta^*$ can be written as a linear combination of the θ-functionals $\theta(L)$, $L \in \Lambda$. □

7. Q-subsets of loops. Let $\mathscr{L}$ be the vector space spanned by $\{\theta(L) : L \in \Lambda\}$, where Λ is the set of all loops in $\bar{M}$. By Proposition 6.2 we always have $\mathscr{L} \subset \mathscr{E}_\theta^*$; and when the hypotheses of Lemma 6.3 are satisfied, we have $\mathscr{L} = \mathscr{E}_\theta^*$. In this section we show that it is not necessary to find all the loops in $\bar{M}$ in order to obtain a spanning set for $\mathscr{L}$.

DEFINITION 7.1. A *Q-subset* of Λ is any subset Φ which can be formed by selecting loops from Λ in the following manner:

> Look for a loop in $\bar{M}$. If one is found, put it in Φ. Select a pair, say (f, g), in this loop and change the (f, g) entry of $\bar{M}$ to 0. Call this new matrix $\bar{M}$ and proceed as before. Stop when no more loops can be found.

For technical reasons we introduce quasi-loops. A *quasi-loop* in $\bar{M}$ is a sequence $K = (f_1, g_1)(f_2, g_2) \cdots (f_n, g_n)$ of an even number n $(n > 0)$ of pairs satisfying all the conditions of Definition 6.1 except possibly condition (i). Let $\theta(K)$ denote the alternating sum of cell expectations associated with the pairs in K (see the paragraph following Definition 6.1).

LEMMA 7.2. *For all quasi-loops K in $\bar{M}$, $\theta(K) \in \mathscr{L}$.*

PROOF. The proof will use induction on the number n of pairs in K. If $n = 2$ then the definition of quasi-loop requires that the two pairs are the same, and so $\theta(K) = 0 \in \mathscr{L}$. Now suppose $n > 2$. If all pairs in K are distinct, then K is a loop and we are done. If any two consecutive pairs are the same, then they can be removed from K leaving a quasi-loop K_0 with $\theta(K_0) = \theta(K)$. By induction $\theta(K_0) \in \mathscr{L}$ and we are done. Hence we can suppose that K has no two consecutive pairs the same but that $(f_i, g_i) = (f_j, g_j)$ for some i and j such that $j \geqq i + 2$. If $i = 1$ and $j = n$, then $K_0 = (f_{n-1}, g_{n-1})(f_2, g_2) \cdots (f_{n-2}, g_{n-2})$ is a quasi-loop with $\theta(K) = -\theta(K_0)$, and again we are done. Otherwise, let $K_1 = (f_p, g_p) \cdots (f_q, g_q)$ where $p = i$ or $i + 1$, whichever is odd, and $q = j$ or $j - 1$, whichever is even. Then K_1 is a quasi-loop with fewer than n pairs and the removal of K_1 from K leaves a quasi-loop K_2. Note that $\theta(K) = \theta(K_1) + \theta(K_2)$. By induction $\theta(K_1)$ and $\theta(K_2)$ are in $\mathscr{L}$, and so $\theta(K) \in \mathscr{L}$. □

LEMMA 7.3. *For any Q-subset Φ of Λ, $\mathscr{L}$ is spanned by $\{\theta(L) : L \in \Phi\}$.*

PROOF. We will argue by induction on the number of nonzero entries in $\bar{M}$. Our induction statement is: If $\bar{M}$ has m nonzero entries and Φ is any Q-subset of Λ, then $\mathscr{L}$ is spanned by $\{\theta(L) : L \in \Phi\}$. This is clear for $m = 0$. Let us assume it is true for all nonnegative integers less than a particular value of m and try to prove it for the integer m.

Take any Q-subset Φ of Λ. Given any $L_1 \in \Lambda$, we must show that $\theta(L_1)$ is a linear combination of θ-functionals $\theta(L)$, $L \in \Phi$. During the selection of the Q-subset Φ it eventually happens that some loop $\tilde{L}$ is selected such that the pair $(\tilde{f}, \tilde{g})$ in $\tilde{L}$ corresponding to the entry of $\bar{M}$ which is changed to 0 is also a pair occurring in L_1. Suppose $\tilde{L}$ is the first such loop in Φ.

Let $\bar{M}^0$ be the matrix which results from changing certain entries (including the $(\tilde{f}, \tilde{g})$ entry) of $\bar{M}$ to 0 according to the portion of the selection procedure which has been completed at the point when $\tilde{L}$ is selected; let Λ^0 be the set of loops in $\bar{M}^0$; let $\mathscr{L}^0$ be the vector space spanned by $\{\theta(L): L \in \Lambda^0\}$; and let Φ^0 be those loops of Φ which are found after $\tilde{L}$ in the selection procedure, i.e., the loops of Φ which are in $\bar{M}^0$. Note that Φ^0 is a Q-subset of Λ^0. Since $\bar{M}^0$ has fewer than m nonzero entries, our inductive assumption allows us to conclude that $\mathscr{L}^0$ is spanned by $\{\theta(L): L \in \Phi^0\}$.

Observe that all the pairs occurring in L_1 and $\tilde{L}$, except the pair $(\tilde{f}, \tilde{g})$, correspond to nonzero entries in $\bar{M}^0$. We can permute the pairs in the loops L_1 and $\tilde{L}$ to get loops $L_1{}^p$ and $\tilde{L}^p$ with $(\tilde{f}, \tilde{g})$ occurring last and first, respectively, and with $\theta(L_1{}^p) = \pm\theta(L_1)$ and $\theta(\tilde{L}^p) = \pm\theta(\tilde{L})$. (To see this it helps to draw a rectilinear picture of a loop.) Form a sequence K of pairs by putting $L_1{}^p$ in front of $\tilde{L}^p$ and removing $(\tilde{f}, \tilde{g})(\tilde{f}, \tilde{g})$. Note that K is a quasi-loop in $\bar{M}^0$ and that $\theta(K) = \theta(L_1{}^p) + \theta(\tilde{L}^p)$. By Lemma 7.2, $\theta(K) \in \mathscr{L}^0$, which means it is a linear combination of θ-functionals $\theta(L)$, $L \in \Phi^0 \subset \Phi$. Now write $\theta(L_1{}^p) = \theta(K) - \theta(\tilde{L}^p)$ and recall that $\theta(\tilde{L}^p) = \pm\theta(\tilde{L})$ and $\tilde{L} \in \Phi$. □

THEOREM 7.4. *Suppose the vectors in F are linearly independent and the vectors in G are linearly independent. Let Φ be any Q-subset of Λ. Then $\mathscr{E}_\theta$ is spanned by D and $\{\theta(L): L \in \Phi\}$.*

PROOF. By Proposition 3.8, $\mathscr{E}_\theta = \mathscr{D} + \mathscr{E}_\theta{}^*$ where $\mathscr{D}$ is the vector space spanned by D. By Lemma 6.3, $\mathscr{E}_\theta{}^* = \mathscr{L}$; and by Lemma 7.3, $\mathscr{L}$ is spanned by $\{\theta(L): L \in \Phi\}$. □

With regard to the size of the spanning set for $\mathscr{E}_\theta$ presented in Theorem 7.4, note that the number of elements in D is equal to the number of elements in T minus the number of equivalence classes in T. It is an interesting fact that the number of loops in Φ is the same for all Q-subsets Φ of Λ (see Example 3 in Section 10). This number is never greater than $(a - 1)(b - 1)$ where a and b are the numbers of elements in F and G respectively.

8. The Q-process. The Q-process is an algorithm for selecting a Q-subset of loops in $\bar{M}$. To apply it we want to put the elements of the index sets F and G into some fixed order. In fact it simplifies the notation if we identify the sets F and G with the sets $\{1, \cdots, a\}$ and $\{1, \cdots b\}$ where a and b are the numbers of elements in F and G respectively.

First, let us describe the *X-process*. The X-process applied to a two-dimensional matrix A consists of the following procedure:

> Look for an entry of A which is the only nonzero entry in its row or in its column. If such an entry is found, change it to 0, call the new matrix A, and proceed as before. The X-process is completed when an A is obtained with no row nor column having only one nonzero entry.

We now describe the *Q-process*:

1) Set $p = 2$.

2) Put the submatrix A of $\bar{M}$ consisting of the first p columns into a temporary working area.

3) Apply the X-process to A.

4) If the pth column of A consists entirely of zeros, go to step (9).

5) Select a position in column p, say (i_1, p), in which a nonzero symbol occurs. Select another position (i_1, j_1), $j_1 \neq p$, in which a nonzero symbol occurs. Select a position (i_2, j_1), $i_2 \neq i_1$, in which a nonzero symbol occurs. Select a position (i_2, j_2), $j_2 \neq j_1$, in which a nonzero symbol occurs. In this manner, alternately look along rows and columns for another nonzero entry until another one is found in column p, say in position (i_r, j_r), $j_r = p$.

6) Put the loop $L = (i_1, p)(i_1, j_1)(i_2, j_1)(i_2, j_2) \cdots (i_r, j_{r-1})(i_r, p)$ in the set Φ.

7) Change the entry in position (i_r, p) to 0 both in A and permanently in $\bar{M}$.

8) Go to step (3).

9) Increase p by 1. If $p \leqq b$, go to step (2). If $p > b$, the Q-process has been completed.

The Q-process eventually terminates because after each loop is found one of the nonzero entries of $\bar{M}$ is changed to 0. In step (3) it is clear that an entry in A is changed to 0 by the X-process only when it cannot be involved in any more loops in A. The sequential nature of the process assures us that at step (4) there are now no loops in the first $p - 1$ columns of $\bar{M}$; therefore if the pth column of A is 0, then there are no loops in A. The procedure of step (5) is possible because the X-process has been applied. Eventually some column must be repeated among the positions selected in step (5) and this column must be column p because there are no loops in the first $p - 1$ columns of $\bar{M}$. It is because of this same fact that we know the sequence L is actually a loop.

In Section 10 three examples are given which involve the Q-process. Perhaps it is best illustrated in the third example.

9. Constraints, degrees of freedom, and reparametrizations. In this section we collect some facts about partitioned linear models which are useful supplementary material to the previous sections. We first summarize some known facts (see Corollary 2.1 in Goldman and Zelen [6], Section 4 in Zyskind [12], Example 2 in Seely [9], and Section 7.1 in Rao and Mitra [7]) about estimable linear parametric functions when there are known linear homogeneous constraints on the parameters. Then in (9.5)—(9.8) we give the generalizations to a partitioned linear model structure. The notation $R(A)$, $N(A)$, and $r(A)$ is used for the range, null space, and rank respectively of a matrix A.

We suppose Y is an $n \times 1$ random vector whose covariance matrix is $\sigma^2 I$ and whose expectation may be written as

$$(9.1) \qquad E(Y) = H\xi + T\theta\,, \qquad \Lambda'\xi = 0\,, \qquad \Gamma'\theta = 0\,.$$

For notational purposes it will sometimes be convenient to disregard the partitioned nature of $E(Y)$ and simply think of $E(Y)$ as

$$(9.2) \qquad E(Y) = X\pi\,, \qquad \Delta'\pi = 0\,,$$

where X, π and Δ are defined in the obvious way, e.g., $X = (H, T)$. For this model a linear parametric function $u'\pi = g'\xi + a'\theta$ is a linear functional on the vector space $N(\Delta')$ of possible parameter vectors. And a linear parametric function $u'\pi$ is said to be estimable provided there exists a linear unbiased estimator for $u'\pi$; or equivalently,

(9.3) $u'\pi$ is estimable if and only if there is some $v \in R(X')$ such that $u'\pi = v'\pi$ for all $\pi \in N(\Delta')$.

Let $\mathscr{E}$ and $\mathscr{E}_\theta$ denote the vector spaces of estimable linear parametric functions of the form $u'\pi$ and $a'\theta = 0'\xi + a'\theta$ respectively; and let $d = \dim \mathscr{E}$ (= degrees of freedom for regression), let $d_\theta = \dim \mathscr{E}_\theta$ (= degrees of freedom for the θ-effects adjusted for the ξ-effects) and let $\bar{d}_\xi$ = degrees of freedom for regression for the submodel $E(Y) = H\xi$, $\Lambda'\xi = 0$. Sections 3–8 are concerned with $\mathscr{E}$ and $\mathscr{E}_\theta$ when there are no restrictions on the model. For determining when the results in Sections 3–8 are applicable to model (9.1) we introduce the conditions

$$\text{C1:} \quad R(X') \cap R(\Delta) = \{0\}\,,$$
$$\text{C2:} \quad R(H') \cap R(\Lambda) = \{0\}\,.$$

Condition C1 says, in the usual terminology, that in model (9.2) the constraints $\Delta'\pi = 0$ are "nonestimable"; whereas condition C2 says that in the submodel $E(Y) = H\xi$, $\Lambda'\xi = 0$, the constraints $\Lambda'\xi = 0$ are "nonestimable". We mention that a classification model with any of the "usual constraints" imposed as restrictions on the parameters will satisfy C1; but C2 will be satisfied only for particular partitions of the parameters. Alternative conditions equivalent to C1 and C2 are, respectively, $d = r(X)$ and $\bar{d}_\xi = r(H)$.

REMARK 9.4. The condition $\Delta = 0$ is equivalent to saying that the parameters in model (9.1) are unrestricted. Thus the statements in this section which apply to the model considered in Sections 3–8 are those we get by assuming $\Delta = 0$. Note that $\Delta = 0$ trivially implies that both C1 and C2 are true.

Concerning linear parametric functions of the form $a'\theta$, (9.3) can be used to establish that

(9.5) $a'\theta$ is estimable if and only if there exists $b \in \{T'z: H'z \in R(\Lambda)\}$ such that $a'\theta = b'\theta$ for all $\theta \in N(\Gamma')$.

Since C2 implies $\{T'z: H'z \in R(\Lambda)\} = \{T'z: H'z = 0\}$, we get

(9.6) If $a_1'\theta, \cdots, a_m'\theta$ would constitute a spanning set for $\mathscr{E}_\theta$ under the assumption $\Delta = 0$, then $a_1'\theta, \cdots, a_m'\theta$ is a spanning set for $\mathscr{E}_\theta$ whenever C2 is true.

It can be established that

$$d = \bar{d}_\xi + d_\theta . \tag{9.7}$$

Furthermore,

(9.8) Suppose C1 and C2 are satisfied. Then

a) If $a_1'\theta, \cdots, a_m'\theta$ would constitute a basis for $\mathscr{E}_\theta$ under the assumption $\Delta = 0$, then $a_1'\theta, \cdots, a_m'\theta$ is a basis for $\mathscr{E}_\theta$.

b) If we set $A = (a_1, \cdots, a_m)$, where the a_i's are as in part (a), and set $W = TA$, then the regression space $\{X\pi: \Delta'\pi = 0\}$ may be expressed as the direct sum $R(H) \oplus R(W)$ and W has full column rank.

We see that the methods of Sections 3–8 will be most useful for model (9.1) when both C1 and C2 are true, because then statement (9.8) applies. Statement (9.8.a) says that we can disregard the constraints in finding a basis for $\mathscr{E}_\theta$ and statement (9.8.b) indicates one use of such a basis. To be more specific, suppose we have the matrix A of (9.8.b). In particular we have $d_\theta = m$. The next step would be to obtain a basis $b_1'\xi, \cdots, b_k'\xi$ for the vector space of estimable linear parametric functions for the submodel $E(Y) = H\xi$, $\Lambda'\xi = 0$. In determining this basis, another application of (9.8.a) tells us we can assume $\Lambda = 0$. Let us suppose we have $B = (b_1, \cdots, b_k)$ where the b_i's are obtained in this way. (It is possible that a suitable partition of the submodel will allow the methods of Sections 3–8 to be utilized.) Of course $\bar{d}_\xi = k$. From (9.7) d is known and we can check for maximal rank. Even if the design matrix does not have maximal rank, the matrices A and B still provide us with an unrestricted full rank reparametrization

$$E(Y) = U\eta + W\psi \tag{9.9}$$

where $U = HB$ and $W = TA$. In other words, $\{X\pi: \Delta'\pi = 0\} = R(U) + R(W)$ and (U, W) has full column rank. Furthermore, $E(Y) = U\eta$ is an unrestricted full rank reparametrization for the submodel $E(Y) = H\xi$, $\Lambda'\xi = 0$. The sums of squares of regression for models (9.1) and (9.9) are the same, and the sum of squares for the θ-effects adjusted for the ξ-effects is equal to the sum of squares for the ψ-effects adjusted for the η-effects. If $(\tilde{\eta}', \tilde{\psi}')'$ is the least-squares estimator for (9.9), then $\tilde{\pi} = (\tilde{\xi}', \tilde{\theta}')'$, where $\tilde{\xi} = B\tilde{\eta}$ and $\tilde{\theta} = A\tilde{\psi}$, satisfies $X'X\tilde{\pi} = X'Y$. Perhaps $\Delta'\tilde{\pi} \neq 0$, but $\tilde{\pi}$ can be converted (see Section 5.7 in [8]) to a least-squares estimator for (9.1).

10. Examples.

EXAMPLE 1. Consider a collection of random variables $\{Y_u\}$ following an additive three-way classification model. The index set for this collection would be $U = \{(i, j, k, p): i = 1, \cdots, a; j = 1, \cdots, b; k = 1, \cdots, c; p = 1, \cdots, n_{ijk}\}$ (if $n_{ijk} = 0$ then no index with the first three components i, j, k occurs in U), so that

$$E(Y_u) = \mu + \alpha_i + \beta_j + \gamma_k, \qquad \text{for} \quad u = (i, j, k, p) \in U.$$

Suppose we are interested in the estimability of γ-contrasts. To apply the procedure of Section 3 we would set $\xi = (\mu, \alpha_1, \cdots, \alpha_a, \beta_1, \cdots, \beta_b)'$ and $\theta = \gamma = (\gamma_1, \cdots, \gamma_c)'$. Then for $u = (i, j, k, p) \in U$, t_u would be the $c \times 1$ vector e_k with its kth component 1 and all other components 0. To apply the procedure of Sections 6–8 we would set $\phi = (\mu, \alpha_1, \cdots, \alpha_a)'$ and $\omega = (\beta_1, \cdots, \beta_b)'$. Then for $u = (i, j, k, p) \in U$, f_u would be the $(1 + a) \times 1$ vector with its first and $(1 + i)$th components 1 and all others 0 and g_u would be the $b \times 1$ vector with its jth component 1 and all others 0. Note that the hypotheses of Theorem 7.4 are satisfied; hence $\mathscr{E}_\gamma$ is spanned by D together with the γ-contrasts derived from a Q-subset of loops. Remark 3.10 also pertains, and so if a basis for $\mathscr{E}_\gamma^*$ is extracted from the spanning set obtained from a Q-subset, then these γ-contrasts together with D constitute a basis for $\mathscr{E}_\gamma$.

Let us consider the specific example where $a = 4$, $b = 3$, $c = 4$, and the occupied cells of the $4 \times 3 \times 4$ incidence matrix $N = (n_{ijk})$ are represented by 1's in the following configuration:

$$\begin{array}{c} \\ \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{array}
\begin{array}{c} \gamma_1 \\ \begin{array}{ccc} \beta_1 & \beta_2 & \beta_3 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & x & x \\ 0 & x & 1 \\ 1 & 0 & 0 \end{bmatrix} \end{array}
\begin{array}{c} \gamma_2 \\ \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & x \\ 0 & 1 & x \\ x & 0 & 0 \end{bmatrix} \end{array}
\begin{array}{c} \gamma_3 \\ \\ \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \end{array}
\begin{array}{c} \gamma_4 \\ \\ \begin{bmatrix} x & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & x \\ x & 0 & 0 \end{bmatrix} \end{array}.$$

Application of the R-process as in Section 3 tells us that the cells marked above by x's are estimable and that $e_1 \sim e_2 \sim e_4$. (The x's above can quickly be obtained via Remark 5.3; and, similarly, we also see directly from the above configuration that $e_1 \sim e_2 \sim e_4$ because the first, second and fourth submatrices are identical and nonzero.) Thus, we can choose $S = \{e_1, e_3\}$ and obtain $D = \{\gamma_2 - \gamma_1, \gamma_4 - \gamma_1\}$.

Referring to Section 6, we now partition ξ into the two subvectors ϕ and ω.

As was done in Section 4, for indexing purposes we can use i in place of f_u and j in place of g_u for $u = (i, j, k, p) \in U$. If we let $\Omega_{e_k} = k$ for $k = 1, 3$, then

$$\bar{M} = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 3 & 0 \end{bmatrix} .$$

Application of the Q-process yields two loops from which it is found that $\mathscr{E}_\gamma{}^*$ is spanned by $\gamma_1 - \gamma_3$. We can conclude that $\mathscr{E}_\gamma$ is spanned by $\{\gamma_2 - \gamma_1, \gamma_4 - \gamma_1, \gamma_1 - \gamma_3\}$; hence all γ-contrasts are estimable.

For the additive three-way model, once a basis for $\mathscr{E}_\gamma$ has been obtained, and hence $d_\gamma = \dim \mathscr{E}_\gamma$ is known, it is a simple matter to obtain $d = \dim \mathscr{E}$. From (9.7) we have that $d = \bar{d}_\xi + d_\gamma$ where $\bar{d}_\xi$ is the degrees of freedom for regression for the submodel

$$E(Y_u) = \mu + \alpha_i + \beta_j , \qquad \text{for} \quad u = (i, j, k, p) \in U .$$

Since this is an additive two-way model, the methods of Section 4 may be applied to calculate $\bar{d}_\xi$. To illustrate, consider the specific example above. Note that $d_\gamma = 3$. The incidence matrix for the two-way submodel is

$$N_{\alpha\beta} = (n_{ij\cdot}) = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 2 & 1 \\ 0 & 2 & 1 \\ 1 & 1 & 0 \end{bmatrix} .$$

Applying the R-process to $N_{\alpha\beta}$ we clearly get a final matrix consisting of all 1's. Thus, from Section 4 we have $\bar{d}_\xi = 4 + 2 = 6$ so that $d = 6 + 3 = 9$. Hence the design matrix for this example has maximal rank.

REMARK 10.1. It is a handy fact that instead of forming $N_{\alpha\beta}$ we can simply use $\bar{M}$ with all its nonzero entries changed to 1's.

EXAMPLE 2. Consider the following Graeco–Latin square with two missing cells:

$$(10.2) \qquad \begin{array}{c} \\ r_1 \\ r_2 \\ r_3 \\ r_4 \end{array} \begin{array}{c} \begin{array}{cccc} c_1 & c_2 & c_3 & c_4 \end{array} \\ \begin{bmatrix} - & B\beta & C\gamma & D\delta \\ B\gamma & A\delta & D\alpha & C\beta \\ C\delta & - & A\beta & B\alpha \\ D\beta & C\alpha & B\delta & A\gamma \end{bmatrix} \end{array} .$$

This corresponds to a collection of 14 random variables following an additive four-way classification model. An appropriate index set would be the set U consisting of the 4-tuples $(1, 2, B, \beta), \cdots (4, 4, A, \gamma)$. For example, $u = (2, 3, D, \alpha) \in U$ and

$$E(Y_u) = \mu + r_2 + c_3 + D + \alpha .$$

Let $\xi = (\mu, r_1, \cdots, r_4, c_1, \cdots, c_4)'$ anc $\theta = (A, B, C, D, \alpha, \beta, \gamma, \delta)'$. The matrix W formed in Section 3 has only one nonzero entry in each column, and the rows in which these nonzero entries appear are different for different columns (this follows from properties of Graeco–Latin squares). Therefore each vector $t \in T$ is equivalent only to itself so that $S = T$ and $\mathscr{D} = \{0\}$. Partition ξ by setting $\phi = (\mu, r_1, \cdots, r_4)'$ and $\omega = (c_1, \cdots, c_4)'$. The hypotheses of Theorem 7.4 are satisfied, and so a spanning set for $\mathscr{E}_\theta$ will be obtained from a Q-subset of loops in $\bar{M}$.

In forming $\bar{M}$ a good choice for the symbol Ω_s is the last two components of the 4-tuple indexing s; thus we may think of the 4×4 array in (10.2) as $\bar{M}$. Applying the Q-process we find a Q-subset consisting of seven loops. From the seven corresponding θ-contrasts we find (after some manipulations) that $d_\theta = 5$ and that a basis for $\mathscr{E}_\theta$ is:

Latin letter contrasts: $B - C,\ A - B - C + D$
Greek letter contrasts: $\beta - \delta,\ \alpha - \beta + \gamma - \delta$
"Inseparable" contrast: $D - A + \alpha - \gamma$.

The Latin letter contrasts above form a basis for the vector space of all estimable Latin letter contrasts. (If not, it would follow that $d_\theta = 6$, which is a contradiction.) Thus, the sum of squares for Latin letter effects adjusted for all other effects leads to a test that $B - C$ and $A - B - C + D$ are zero rather than the usual test that all Latin letter effects are equal. The degrees of freedom associated with this sum of squares is 2.

To calculate d, note that the submodel with parameter vector ξ is an additive two-way model. Then use Remark 10.1 to find $\bar{d}_\xi = 7$, so that $d = 7 + 5 = 12$ by (9.7).

From the above analysis we have d and the relevant information concerning the Greek and Latin letter effects. To obtain similar information for the row and column effects one could interchange the roles of ξ and θ. For certain facts, however, alternative methods are available at this stage. For example, suppose we want the degrees of freedom d_r for the row effects adjusted for all other effects. Apply (9.7) with the parameters partitioned so that $\theta = (r_1, r_2, r_3, r_4)'$. We already know $d = 12$. To get $\bar{d}_\xi$ we must consider the submodel obtained by dropping the θ-effects, which is an additive three-way model with 10 degrees of freedom for regression. Therefore, $d_r = 12 - 10 = 2$.

REMARK 10.3. The choice $\theta = (A, B, C, D, \alpha, \beta, \gamma, \delta)'$ in Example 2 was motivated by the fact that the vector ξ of remaining parameters could be partitioned into ϕ and ω such that the hypotheses of Theorem 7.4 would be satisfied. Even if we were only interested in Latin letter contrasts, this would still be a better choice in Example 2 than $\theta = (A, B, C, D)'$. This latter choice does not allow us to apply Theorem 7.4 and, in fact, leads to no information about estimable Latin letter contrasts. Note, however, that for any partition of the parameter vector into ϕ, ω, and θ one can always apply Proposition 6.2, and in some cases this will lead to all the desired estimability information.

EXAMPLE 3. Consider a collection of random variables $\{Y_u\}$ following a two-way classification model with interaction. The index set is $U = \{(i, j, k) : i = 1, \cdots, a; j = 1, \cdots, b; k = 1, \cdots, n_{ij}\}$ where $n_{ij} = 0$ implies there is no $u \in U$ whose first two components are i, j. Thus, our model is

$$E(Y_u) = \mu + \alpha_i + \beta_j + \theta_{ij}, \qquad \text{for} \quad u = (i, j, k) \in U .$$

Suppose there are no restrictions on the parameters. Then it is known that there are no estimable α-contrasts or β-contrasts. So we concentrate on the estimable θ-contrasts. Partition the model according to $\xi = (\mu, \alpha_1, \cdots, \alpha_a, \beta_1, \cdots, \beta_b)'$ and θ equal to the vector of θ_{ij}'s ordered lexicographically. For $u = (i, j, k) \in U$ let $e_{ij} = t_u$ so that $e_{ij} \cdot \theta = \theta_{ij}$ and $T = \{e_{ij} : n_{ij} \neq 0\}$. Let ϕ, ω, f_u and g_u be defined as in Example 1. We note that the hypotheses of Theorem 7.4 are satisfied and hence $\mathscr{E}_\theta$ is spanned by D together with the θ-contrasts derived from a Q-subset of loops in $\bar{M}$.

The column indexed by e_{ij} in the matrix W (to which the R-process is applied to obtain D) has a nonzero entry only in the row indexed by the vector h satisfying $h \cdot \xi = \mu + \alpha_i + \beta_j$. Therefore, $S = T$ and $\mathscr{D} = \{0\}$. In forming $\bar{M}$ we can choose Ω_s to be the numeral 1 for all $s \in S$ because the position of a 1 in $\bar{M}$ tells which $s \in S$ is associated with it. Hence we can obtain $\bar{M}$ from the incidence matrix $N = (n_{ij})$ by changing all nonzero entries to 1.

Suppose a Q-subset of loops in $\bar{M}$ is selected. Then each θ-contrast $\theta(L)$ derived from a loop L in the Q-subset contains a term θ_{ij} which does not occur in any contrasts derived from loops selected later. Therefore, these θ-contrasts not only form a spanning set for $\mathscr{E}_\theta$ but in fact form a basis. (This leads to the fact mentioned in Section 7 that all Q-subsets contain the same number of loops.)

REMARK 10.4. For the two-way model with interaction it is easily seen that d is equal to the number of nonzero entries in the incidence matrix N. By (9.7) we have $d = \bar{d}_\xi + d_\theta$. Thus, an alternative method to that in Section 4 for calculating the rank of the design matrix for an additive two-way classification model with incidence matrix N is to subtract the number of loops in a Q-subset of loops in N from the number of nonzero entries in N.

As a specific case of Example 3 suppose $a = 4$, $b = 5$, and that the incidence matrix $N = (n_{ij})$ is

$$N = \begin{bmatrix} 3 & 2 & 2 & 0 & 0 \\ 0 & 1 & 4 & 2 & 0 \\ 2 & 0 & 1 & 0 & 0 \\ 1 & 3 & 0 & 1 & 2 \end{bmatrix}, \quad \text{and so} \quad \bar{M} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \end{bmatrix}.$$

The Q-process applied to $\bar{M}$ yields four loops with corresponding contrasts

$$\theta_{12} - \theta_{11} + \theta_{41} - \theta_{42}, \qquad \theta_{13} - \theta_{12} + \theta_{22} - \theta_{23},$$
$$\theta_{13} - \theta_{11} + \theta_{31} - \theta_{33}, \qquad \theta_{24} - \theta_{22} + \theta_{12} - \theta_{11} + \theta_{41} - \theta_{44}.$$

Hence $d_\theta = 4$. Also, from Remark 10.4 we have $d = 12$.

REMARK 10.5. In Example 3 we have assumed no constraints on the parameters. Write the model in the partitioned form of (9.1) as $E(Y) = H\xi + T\theta$ where ξ and θ are as in Example 3 and suppose we have any of the "usual" constraints, say $\Lambda'\xi = 0$ and $\Gamma'\theta = 0$, on the parameters. Then C1 and C2 in Section 9 are true so that a basis for $\mathscr{E}_\theta$ can be obtained from the θ-contrasts derived from the loops in a Q-subset. However, unlike additive classification models, statement (9.8) will not apply if the roles of θ and ξ are interchanged because $R(T') \cap R(\Gamma) \neq \{0\}$.

Acknowledgments. The authors wish to thank the referees and the associate editor for their helpful comments.

REFERENCES

[1] BIRKES, D., DODGE, Y., HARTMANN, N., and SEELY, J. (1972). Estimability and analysis for additive two-way classification models with missing observations. Technical Report No. 30, Department of Statistics, Oregon State Univ.

[2] BIRKES, D., DODGE, Y., and SEELY J. (1972). Estimability and analysis for additive three-way classification models with missing observations. Technical Report No. 33, Department of Statistics, Oregon State Univ.

[3] BOSE, R. C. (1949). Least square aspects of analysis of variance. Inst. Stat., Mimeo Series 9, Chapel Hill, N.C.

[4] DODGE, Y. (1974). Estimability considerations of N-way classification experimental arrangements with missing observations. Ph. D. dissertation, Oregon State Univ.

[5] ECCLESTON, J. A. and HEDAYAT, A. (1974). On the theory of connected designs: characterization and optimality. *Ann. Statist.* **2** 1238–1255.

[6] GOLDMAN, A. J. and ZELEN, M. (1964). Weak generalized inverses and minimum variance linear unbiased estimation. *J. Res. Nat. Bur. Stand.* **62B** 151–172.

[7] RAO, C. R. and MITRA, S. K. (1971). *Generalized Inverse of Matrices and Its Applications.* Wiley, New York.

[8] SEARLE, S. R. (1971). *Linear Models.* Wiley, New York.

[9] SEELY, J. (1970). Linear spaces and unbiased estimation. *Ann. Math. Statist.* **41** 1725–1734.

[10] SRIVASTAVA, J. N. and ANDERSON, D. A. (1970). Some basic properties of multidimensional partially balanced designs. *Ann. Math. Statist.* **41** 1438–1445.

[11] WEEKS, D. L. and WILLIAMS, D. R. (1964). A note on the determination of connectedness in an N-way cross classification. *Technometrics* **6** 319–324. Errata, *Technometrics* **7** 281.

[12] ZYSKIND, G. (1967). On canonical forms, nonnegative covariance matrices and best and simple least squares linear estimators in linear models. *Ann. Math. Statist.* **38** 1092–1109.

DEPARTMENT OF STATISTICS
OREGON STATE UNIVERSITY
CORVALLIS, OREGON 97331