Survey Methodology, June 1994
Vol. 20, No. 1, pp. 65–73
Statistics Canada

Multi-way Stratification by Linear Programming

R.R. SITTER and C.J. SKINNER[1]

ABSTRACT

Rao and Nigam (1990, 1992) showed how a class of controlled sampling designs can be implemented using linear programming. In this article their approach is applied to multi-way stratification. A comparison is made with existing methods both by illustrating the sampling schemes generated for specific examples and by evaluating mean squared errors. The proposed approach is relatively simple to use and appears to have reasonable mean squared error properties. The computations required can, however, increase rapidly as the number of cells in the multi-way classification increase. Variance estimation is also considered.

KEY WORDS: Controlled selection; Linear programming; Multistage sampling; Stratified sampling.

1. INTRODUCTION

There are often several stratifying variables available to the sample designer and it is natural in such cases for the designer to consider defining strata as the cells formed by cross-classifying categories of these variables. A problem with this approach, particularly common when selecting primary sampling units (psu's) in household surveys, is that the desired sample size may be less than the total number of cells and hence conventional methods of stratification may be inapplicable.

An illustration, based on a hypothetical example of Bryant *et al.* (1960), is given in Table 1. Communities (psu's) are classified by two stratifying factors: type of community with three categories and region with five categories. The desired sample size of $n = 10$ is less than the total number of cells, 15. This example also illustrates a related problem. The entries in Table 1 are the expected counts under proportionate stratification, that is the population proportions multiplied by the sample size. Even if the sample size was doubled to exceed the number of cells, the expected sample counts would still not be integers. Whilst the effect of rounding such values to integers may not be practically significant for large expected counts, the choice of how to round with very small expected counts may be of greater concern.

Table 1

Expected Sample Cell Counts Under Proportionate Stratification with $n = 10$

Regions	Type of Community			
	Urban	Rural	Metropolitan	Total
1	1.0	0.5	0.5	2.0
2	0.2	0.3	0.5	1.0
3	0.2	0.6	1.2	2.0
4	0.6	1.8	0.6	3.0
5	1.0	0.8	0.2	2.0
Total	3.0	4.0	3.0	10.0

One reaction to the problem of many cells is simply to drop one or more of the stratifying variables or to group some of the categories. Alternatively, a number of procedures have been proposed which attempt to retain some 'control' for all the categories of all the stratifying variables by permitting different forms of random selection of cells.

Goodman and Kish (1950) proposed one procedure under the title 'controlled selection'. Jessen (1970) suggests that 'this method is somewhat complicated and its use in applied sampling appears limited' (p. 778). Waterton (1983) illustrates this complexity. Bryant *et al.* (1960) propose a much simpler method for two-way stratification. Their method has the property that the expected sample counts display independence between the rows and columns of the two-way table. If the rows and columns are also independent in the population then there is no problem but if, as will often be the case, there is an appreciable lack of independence then some reweighting will usually be necessary and this can be unattractive in practice and can inflate the variance as is shown in Section 5. Jessen (1970) points out that a further limitation of the method of Bryant *et al.* (1960) is that it is not possible to constrain specified cell sizes to be zero. He proposes two approaches for both two-way and three-way stratification but both approaches remain fairly complicated to implement and, as noted by Causey *et al.* (1985), do not always lead to a solution.

All the above methods may be carried out by hand with varying degrees of laboriousness, but none take advantage of the power of modern computing. In this paper we shall show how computational procedures of linear programming can be applied to the multi-way stratification problem following Rao and Nigam (1990, 1992). Our proposed approach may be viewed as complementing the linear

[1] R.R. Sitter, Department of Mathematics and Statistics, Carleton University, Ottawa, Ontario K1S 5B6; C.J. Skinner, Department of Social Statistics, University of Southampton S09 5NH, U.K.

programming approach proposed by Causey *et al.* (1985). Which of the two approaches is preferable will depend on the nature of the stratification problem and on the software available. The potential disadvantage of our approach is that it can be much more computationally intensive, since the number of unknowns in our linear programming problem may be as large as $\binom{k}{n}$, when k is the number of cells in the table and n is the sample size, whereas the number of unknowns in the approach of Causey *et al.* (1985) is only k. A number of suggestions will be made, however, to reduce the computational demands of our approach. There are several potential advantages of our approach. First, the stratification problem corresponds directly to the linear programming problem and so the computer programming is straightforward, whereas the approach of Causey *et al.* is less direct, involving mimicking the behaviour of nonlinear functions by linear functions (p. 904) and nesting repeated linear programming problems within a further recursive algorithm. Second, our procedure always has a solution, whereas the procedure of Causey *et al.* need not, for example in cases of three-way stratification. Third, the objective function in our linear programming problem can be naturally modified to reflect the different objectives of the stratification problem, for example in a three-way problem where it is more important to 'balance' the sample with respect to the first two stratifying variables than the third. Fourth, our procedure can be naturally modified to constrain the joint inclusion probabilities of cells to be positive in order to permit unbiased variance estimation.

2. THE PROPOSED APPROACH

2.1 Basic Ideas

We begin with the simplest kind of two-way stratification. Let a population of N units be classified into the RC cells of a two-way table formed by cross-classifying a row stratification factor with R categories and a column factor with C categories. Let N_{ij} be the number of units in cell ij, that is the set of units in both row i and column j, and let $P_{ij} = N_{ij}/N$ be the corresponding proportion. The parameter of interest is taken to be the population mean, $\bar{Y}$, of a variable Y.

Consider the following two-stage sampling procedure. First, sample sizes n_{ij} are determined for each cell according to a specified randomized procedure. Letting s denote the $R \times C$ array (n_{ij}, $i = 1, \ldots, R, j = 1, \ldots, C$), this procedure assigns a probability $p(s)$ to each s in a set S of possible arrays. To emphasize the dependence of n_{ij} on s we write $n_{ij}(s)$. Second, a simple random sample of $n_{ij}(s)$ units is selected from cell ij and the values of Y are recorded for the sample units.

We restrict attention to designs of fixed sample size $n > 0$, that is we restrict S to be the set S_n of all arrays such that

$$\sum_{i=1}^{R} \sum_{j=1}^{C} n_{ij}(s) = n.$$

We also restrict attention to proportionate stratification so that

$$\sum_{s \in S_n} n_{ij}(s)p(s) = nP_{ij} \quad \text{for} \quad i = 1, \ldots, R, \quad j = 1, \ldots, C. \tag{2.1}$$

It follows from (2.1) that the simple unweighted sample mean $\bar{y}(s)$ is an unbiased estimator of $\bar{Y}$. We propose to choose a (or the) sampling design $p(s)$ which minimizes the expected lack of 'desirability' of the sample s by solving the problem:

$$\underset{p \in P}{\text{minimize}} \sum_{s \in S_n} w(s)p(s), \tag{2.2}$$

subject to the constraint (2.1), where $w(s)$ is a loss function for the sample s to be specified and P is the class of possible sample designs on S_n obeying

$$0 \leq p(s) \leq 1 \quad \text{for all} \quad s \in S_n. \tag{2.3}$$

Note that (2.1) implies $\sum_{s \in S_n} p(s) = 1$. The key observation of Rao and Nigam (1990, 1992) is that the objective function in (2.2) and the equality and inequality constraints in (2.1) and (2.3) are all linear in $p(s)$ and hence this problem may be solved directly by linear programming with the $p(s)$, $s \in S_n$, as unknowns. The main obstacle to this approach is that the number of elements in S_n is often very large and even with modern computing power it becomes difficult to carry out linear programming if the number of unknowns is large.

It is therefore desirable to restrict attention to a subset of S_n. One natural restriction is to consider only arrays s for which $n_{ij}(s)$ is either equal to $I_{ij} = [nP_{ij}]$, the greatest integer less than nP_{ij}, or $I_{ij} + 1$. Letting $\tilde{n}_{ij}(s) = n_{ij}(s) - I_{ij}$ and $r_{ij} = nP_{ij} - I_{ij}$ the problem becomes

$$\underset{p \in P}{\text{minimize}} \sum_{s \in \tilde{S}_{\tilde{n}}} w(s)p(s), \tag{2.4}$$

subject to

$$\sum_{s \in \tilde{S}_{\tilde{n}}} \tilde{n}_{ij}(s)p(s) = r_{ij}, \tag{2.5}$$

$$\sum_{s \in \tilde{S}_{\tilde{n}}} p(s) = 1, \; 0 \leq p(s) \leq 1 \quad \text{for all} \quad s \in \tilde{S}_{\tilde{n}}, \tag{2.6}$$

where $\tilde{S}_{\tilde{n}}$ is the set of $R \times C$ arrays, where all elements are 0 or 1 and the sum of elements is $\tilde{n} = n - \sum_{ij} I_{ij}$. Note, of course, that if all the I_{ij} are zero, then this is just the same problem as before. The number of elements in $\tilde{S}_{\tilde{n}}$, which determines the magnitude of the computational task for linear programming, is now $\binom{RC}{\tilde{n}}$. This number can still be very large, however, and some further reduction can be achieved by sensible choice of the loss function $w(s)$ as discussed in the next section.

For Table 1, this would amount to considering the situation represented by Table 2, while only allowing a 0 or 1 cell sample size, and then adding back 1 to cells (1,1), (3,3), (4,2) and (5,1) in the final solution. Thus $n = 10$, but $\tilde{n} = 6$.

Table 2

Table of r_{ij} Values from Table 1 with $\tilde{n} = 6$

Regions	Type of Community			
	Urban	Rural	Metropolitan	Total
1	0.0	0.5	0.5	1.0
2	0.2	0.3	0.5	1.0
3	0.2	0.6	0.2	1.0
4	0.6	0.8	0.6	2.0
5	0.0	0.8	0.2	1.0
Total	1.0	3.0	2.0	6.0

2.2 Choice of Loss Function *w(s)*

The major flexibility of the proposed approach derives from the user's freedom to choose the function $w(s)$ which enters the objective function in (2.2). The conventional approach to two-way stratification (*e.g.*, Jessen 1970; Causey *et al.* 1985) is to require that the selected sample s obey the marginal constraints:

$$| n_{i\cdot}(s) - nP_{i\cdot} | < 1 \quad i = 1, \ldots, R, \tag{2.7}$$

$$| n_{\cdot j}(s) - nP_{\cdot j} | < 1 \quad j = 1, \ldots, C, \tag{2.8}$$

where

$$n_{i\cdot}(s) = \sum_j n_{ij}(s), \quad n_{\cdot j}(s) = \sum_i n_{ij}(s)$$

$$P_{i\cdot} = \sum_j P_{ij}, \quad P_{\cdot j} = \sum_i P_{ij}.$$

This requirement can be accommodated in our approach by setting $w(s)$ as (effectively) infinite for samples s not satisfying (2.7) or (2.8) or more simply by excluding such samples from the set S_n. The problem with this conventional approach is that no solution to the constrained optimization-problem may exist.

In our approach, however, if we use a loss function such as

$$w(s) = \sum_{i=1}^{R} (n_{i\cdot}(s) - nP_{i\cdot})^2 + \sum_{j=1}^{C} (n_{\cdot j}(s) - nP_{\cdot j})^2, \tag{2.9}$$

then an optimal solution will always exist within a large enough set S_n. In practice, it may be advantageous computationally to restrict the set S_n initially to only those samples obeying (2.7) and (2.8), or even a subset of these, and then to expand the set if necessary, say by changing 1 to 2 in (2.7) and (2.8), until a solution is found.

Let us now consider the more fundamental question of why constraints such as (2.7) and (2.8) are sensible anyway. From a non-statistical point of view, the balancing of a sample with respect to factors with a known population distribution may reassure users about the 'representativeness' of the sample. From a statistical point of view, given our unbiasedness constraint (2.1), it is natural to consider how the loss function might be chosen to improve efficiency. This question may be examined by taking $w(s)$ as the mean squared error $E_m(\bar{y}(s) - \bar{Y})^2$ under a model m. Then the solution to the optimization problem (2.2) minimizes the design-expected model-mean squared error or equivalently, since we require design-unbiasedness, the model-expected design variance.

Consider, for example, the main-effects analysis of variance model

$$y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk},$$

where y_{ijk} is the kth value of Y in cell ij, μ is a fixed mean and α_i, β_j and ϵ_{ijk} are independent zero-mean random effects with variances σ_α^2, σ_β^2 and σ_ϵ^2, respectively. Then, ignoring finite population correction terms,

$$E_m(\bar{y}(s) - \bar{Y})^2 = \sigma_\alpha^2 \sum_i (n_{i\cdot}(s)/n - P_{i\cdot})^2 + \sigma_\beta^2 \sum_j (n_{\cdot j}(s)/n - P_{\cdot j})^2 + \sigma_\epsilon^2/n. \tag{2.10}$$

Hence, if $\sigma_\alpha^2 = \sigma_\beta^2$ the expected design variance of $\bar{y}(s)$ under this model is minimized by taking the loss function in (2.9). Alternatively, if one had some prior information about the likely ratio of the between row variance relative to the between column variance then it may be sensible, on efficiency grounds, to modify the loss function in (2.9) by multiplying the first term on the right hand side of (2.9) by this estimated ratio.

On the other hand if it is thought *a priori* that there is likely to be a strong interaction between the row and column factors in their effect on Y then simply attempting to balance on the margins may be inappropriate. For example, if one stratification factor is urban/rural and the other is an economic indicator X and it is known that Y is positively related to X in urban areas and negatively related in rural areas then it is likely to be more efficient to stratify partially by X *separately* within rural and urban areas than to balance fully on both margins. See Bryant *et al.* (1960, section 9) for related comments on efficiency for two-way stratification.

2.3 Higher-way Stratification

The proposed approach extends naturally to 3 or more stratifying factors by letting s denote the corresponding r-way array. The loss function will typically include further terms, for example for three-way stratification we might take

$$w(s) = \lambda_1 \sum_{i=1}^{R_1} (n_{i\cdot\cdot}(s) - nP_{i\cdot\cdot})^2 + \lambda_2 \sum_{j=1}^{R_2} (n_{\cdot j\cdot}(s) - nP_{\cdot j\cdot})^2 + \lambda_3 \sum_{k=1}^{R_3} (n_{\cdot\cdot k}(s) - nP_{\cdot\cdot k})^2$$

in obvious notation, where λ_1, λ_2 and λ_3 are included to represent the relative importance of balancing on the three factors and might consist of prior estimates of the variances of the Y means between categories of the three stratifying factors, as in (2.10).

2.4 Multistage Sampling

One important practical application of multi-way stratification is to the selection of primary sampling units (psu's) in multistage sampling, where it is common for information of several stratifying factors to be available.

In the approach of Section 2.1, the inclusion probabilities of each population unit are $E(n_{ij}(s)/N_{ij}) = n/N$. If it is desired to select psu's with equal probability then this approach extends immediately with the psu's constituting the units and with the observed values of Y replaced by unbiased estimators of the psu totals. Suppose instead that it is desired to select psu's with unequal probabilities, say nz_{ijk} for psu k in cell ij, where usually z_{ijk} will equal $M_{ijk}/\sum_{ijk} M_{ijk}$, with M_{ijk} being some measure of size of psu k in cell ij. Then the procedure may be simply modified by setting P_{ij} equal to the sum of z_{ijk} over psu's k in cell ij. Then, if $n_{ij}(s) > 0$, a sample of $n_{ij}(s)$ psu's in cell ij is selected by some probability proportional to z_{ijk} method.

3. EXAMPLES

Example 1: Bryant, Hartley and Jessen (1960)

We will first demonstrate the method on the hypothetical example of Bryant *et al.* (1960) given in Table 1. We first reduce the problem to the form of (2.4), (2.5) and (2.6), where the r_{ij}'s are given in Table 2. The weight function in (2.9) in this reduced linear programming problem becomes

$$w(s) = \sum_{i=1}^{5} (\tilde{n}_{i\cdot}(\tilde{s}) - r_{i\cdot})^2 + \sum_{j=1}^{3} (\tilde{n}_{\cdot j}(\tilde{s}) - r_{\cdot j})^2.$$

Applying a standard linear programming package in the NAG FORTRAN library, we obtain the solution given in Table 3. The I_{ij} values have been added to the solution so that $n_{ij} = I_{ij} + \tilde{n}_{ij}(\tilde{s})$. It turns out for this solution that each s, for which $p(s) > 0$, has margins $n_{i\cdot}(s)$ and $n_{\cdot j}(s)$ which match the desired margins exactly, that is the solution makes (2.4) zero.

Table 3

Solution to Example 1

s	$p(s)$	s	$p(s)$
1 1 0 1 0 0 0 1 1 0 2 1 1 0 1	0.2	1 1 0 0 0 1 0 1 1 1 1 1 1 1 0	0.1
1 1 0 0 0 1 0 0 2 1 2 0 1 1 0	0.2	1 0 1 0 1 0 1 0 1 0 2 1 1 1 0	0.2
1 0 1 0 1 0 0 1 1 1 1 1 1 1 0	0.1	1 0 1 0 0 1 0 1 1 1 2 0 1 1 0	0.2

Example 2: Jessen (1970)

Jessen (1970) proposed two methods for two-way and three-way stratification. Both of these are quite complicated and involve determining the set of samples which exactly match the margins. Neither method is guaranteed to yield a solution. Jessen (1970) applies both methods to a simple hypothetical example for which both yield a solution. This example is reproduced in Table 4. In this example, since all of the $nP_{ij} < 1$, the linear programming problems defined by (2.1), (2.2) and (2.3) and by (2.4), (2.5) and (2.6), respectively, are identical. We applied our method to this problem, again using the $w(s)$ as defined in (2.9). By trying a number of different seeds in the optimization routine, we were able to obtain three different solutions, all of which make (2.2) zero and satisfy the constraints. These are given in Table 5. The first two solutions are the same two as obtained by Jessen's method 2 and method 3, respectively.

Table 4

Example 2: Jessen (1970)
Expected Sample Cell Counts Under Proportionate Stratification with $n = 6$

Rows	Columns 1	2	3	$nP_{i\cdot}$
1	0.8	0.5	0.7	2.0
2	0.7	0.8	0.5	2.0
3	0.5	0.7	0.8	2.0
$nP_{\cdot j}$	2.0	2.0	2.0	6.0

Table 5

Solution to Example 2

s	$p_1(s)$	$p_2(s)$	$p_3(s)$
1 0 1 1 1 0 0 1 1	0.5	0.4	0.3
1 1 0 0 1 1 1 0 1	0.3	0.2	0.1
0 1 1 1 0 1 1 1 0	0.2	0.1	0.0
1 1 0 1 0 1 0 1 1	0.0	0.1	0.2
1 0 1 0 1 1 1 1 0	0.0	0.1	0.2
0 1 1 1 1 0 1 0 1	0.0	0.1	0.2

Example 3: Causey, Cox and Ernst (1985)

Causey *et al.* (1985) give an example of three-way stratification for which their method fails to yield a solution. They consider a population subject to a $2 \times 2 \times 2$ stratification from which a sample of size $n = 2$ is to be drawn, with the expected sample size in the ijk-th cell, n_{ijk}, as follows:

$$n_{111} = n_{221} = n_{122} = n_{212} = .5$$

$$n_{121} = n_{211} = n_{112} = n_{222} = 0.$$

If we apply our method in a similar manner to Examples 1 and 2 we obtain the solution given in Table 6. In this case, the objective function did not attain zero so that the margins are not exactly matched in each sample.

Table 6

Solution to Example 3

s: $i = 1$	$i = 2$	$p(s)$
1 0 0 0	0 1 0 0	0.5
0 0 0 1	0 0 1 0	0.5

4. COMPARISON OF MSE

In this section the mean squared error (MSE) of the proposed design with estimator $\bar{y}$ will be compared with the MSE of the design of Bryant *et al.* (1960) with either of the two estimators they propose, namely $\bar{y}_U$ and $\bar{y}_B$, where the U and B subscripts indicate that the first estimator is unbiased and the second is not. Let the cells be denoted c (ij in the two-way case), let k (and where necessary l) denote a unit within a cell, and suppress the s in $n_c(s)$ for simplicity of notation. The inclusion probability of any unit k in cell c is

$$\pi_{ck} = E[n_c]/N_c = E[n_c]/(NP_c) \tag{4.1}$$

and the joint inclusion probability of unit k in cell c and unit k' in cell c' is

$$\pi_{ckc'k'} = \begin{cases} \dfrac{E[n_c(n_c-1)]}{N_c(N_c-1)} & \text{if} \quad c = c' \\[2ex] \dfrac{E[n_c n_{c'}]}{N_c N_{c'}} & \text{if} \quad c \neq c'. \end{cases} \tag{4.2}$$

For large N this is approximately

$$\pi_{ckc'k'} \doteq \frac{E(n_c n_{c'})}{N^2 P_c P_{c'}} - \frac{E(n_c)}{N^2 P_c^2} I_{[c=c']}, \tag{4.3}$$

where

$$I_{[c=c']} = \begin{cases} 1 & \text{if} \quad c = c' \\ 0 & \text{if} \quad c \neq c'. \end{cases}$$

The expectations will differ for our design compared to the Bryant *et al.* design and thus the π_{ck} and $\pi_{ckc'k'}$ will differ. Keeping this in mind we can obtain the variance of $\bar{y}$, $\bar{y}_U$ and $\bar{y}_B$ in a generalized form in terms of the π_{ck}

and $\pi_{ckc'k'}$ values and thus have some basis for comparison. To do this, let us consider an estimator of the form $\bar{z} = \sum_c \sum_k w_c y_{ck}/n$, where the w_c values are fixed known constants independent of k. If we restrict to the case where $n_{i\cdot} = nP_{i\cdot}$ and $n_{\cdot j} = nP_{\cdot j}$, that is, integer marginal requirements, then both of the estimators given in Bryant *et al.* as well as our estimator are of this form. We will assume this to be the case in the sequel. Replacing the subscript c with ij for two-way stratification, $\bar{y}_U$ and $\bar{y}_B$ are of the same form as $\bar{z}$ with $w_c = w_{ij} = G_{ij} = P_{ij}/(P_{i\cdot}P_{\cdot j})$ and $w_c = w_{ij} = 1$, respectively. The estimator $\bar{y}$ is also of the form $\bar{z}$ with $w_c = w_{ij} = 1$.

We can now obtain a general form for the variance of $\bar{z}$ keeping in mind that the π_{ck} and $\pi_{ckc'k'}$ values will differ for the Bryant *et al.* design and our design:

$$V(\bar{z}) = \frac{1}{2n^2} \sum_c \sum_{c'} \sum_k \sum_{k'} (\pi_{ck}\pi_{c'k'} - \pi_{ckc'k'}) (w_c y_{ck} - w_{c'} y_{c'k'})^2. \quad (4.4)$$

Using (4.1) and (4.3) this becomes

$$V(\bar{z}) = \frac{1}{2n^2} \sum_c \frac{w_c^2 E(n_c)}{N^2 P_c^2} \sum_k \sum_{k'} (y_{ck} - y_{ck'})^2 - \frac{1}{2n^2} \sum_c \sum_{c'} \frac{\mathrm{Cov}(n_c, n_{c'})}{N^2 P_c P_{c'}} \sum_k \sum_{k'} (w_c y_{ck} - w_{c'} y_{c'k'})^2. \quad (4.5)$$

Noting that

$$\sum_k \sum_l (y_{ck} - y_{cl})^2 = 2N^2 P_c^2 S_c^2$$

and

$$\sum_k \sum_{k'} (w_c y_{ck} - w_{c'} y_{c'k'})^2 = N^2 P_c P_{c'} [w_c^2 S_c^2 + w_{c'}^2 S_{c'}^2 + (w_c \bar{Y}_c - w_{c'} \bar{Y}_{c'})^2],$$

where S_c^2 refers to the population variance of cell c, (4.5) reduces to

$$V(\bar{z}) = \frac{1}{n^2} \sum_c w_c^2 E(n_c) S_c^2 - \frac{1}{2n^2} \sum_c \sum_{c'} \mathrm{Cov}(n_c, n_{c'}) [w_c^2 S_c^2 + w_{c'}^2 S_{c'}^2 + (w_c \bar{Y}_c - w_{c'} \bar{Y}_{c'})^2]$$
$$= v_1 + v_2, \quad \text{say}. \quad (4.6)$$

The first term v_1 may be interpreted as the usual stratified variance for fixed sample sizes $E(n_c)$ within the two-way 'strata' (of course in our case the $E(n_c)$ will generally not be integers). The second term v_2 may be interpreted as the increase in variance arising from the variability of the n_c and the correlation between them. We discuss this further at the end of this section. We now revert to the notation $c = ij$ and compare the variances for two-way stratification.

First let us consider v_1 in (4.6). For the Bryant *et al.* method $E(n_{ij}) = nP_{i\cdot}P_{\cdot j}$, $\bar{y}_U = \sum_i \sum_j \sum_k G_{ij} y_{ijk}/n$, $G_{ij} = P_{ij}/(P_{i\cdot}P_{\cdot j})$ and $\bar{y}_B = \sum_i \sum_j \sum_k y_{ijk}/n$.

Thus

$$v_1(\bar{y}_U) = \sum_i \sum_j P_{ij} G_{ij} S_{ij}^2 / n,$$

(this is the same as the first term of equation (12) in Bryant *et al.*) and

$$v_1(\bar{y}_B) = \sum_i \sum_j P_{i\cdot}P_{\cdot j} S_{ij}^2 / n.$$

In the case of our approach $E(n_{ij}) = nP_{ij}$ and $\bar{y} = \sum_i \sum_j \sum_k y_{ijk}/n$ so that

$$v_1(\bar{y}) = \sum_i \sum_j P_{ij} S_{ij}^2 / n.$$

Next let us consider v_2. It is not difficult to show that for both the Bryant *et al.* method and our approach (see Appendix)

$$\sum_i \mathrm{Cov}(n_{ij}, n_{i'j'}) = \sum_j \mathrm{Cov}(n_{ij}, n_{i'j'}) = 0. \quad (4.7)$$

Using this and replacing c and c' by ij and $i'j'$, respectively, in v_2 given in (4.6), it follows that v_2 reduces to

$$v_2 = \frac{1}{n^2} \sum_i \sum_j \sum_{i'} \sum_{j'} \mathrm{Cov}(n_{ij}, n_{i'j'}) w_{ij} w_{i'j'} \bar{Y}_{ij} \bar{Y}_{i'j'}.$$

Replacing w_{ij} with G_{ij} we get $v_2(\bar{y}_U)$, and using simple algebra one can show that this is the same as term 2 of equation (12) in Bryant *et al.* Replacing w_{ij} with 1 gives the form of $V(\bar{y}_B)$ and of $V(\bar{y})$, noting that the $\mathrm{Cov}(n_{ij}, n_{i'j'})$ will not be the same for both. So we see that v_2 depends only on the cell means while v_1 depends only on the within cell variances.

Finally, we should note that

$$\mathrm{bias}(\bar{y}_B) = -\sum_i \sum_j (P_{ij} - P_{i\cdot}P_{\cdot j}) \bar{Y}_{ij}, \quad (4.8)$$

since to compare the three estimators the mean square error (MSE) will be the relevant measure, and this bias will contribute to $\mathrm{MSE}(\bar{y}_B)$.

Combining the expressions for v_1, v_2 and bias$(\bar{y}_B)$ above permits an analytical comparison of the MSE of the proposed approach with that of the approach of Bryant *et al.* (1960) using either $\bar{y}_U$ and $\bar{y}_B$. It is difficult, however, to make general statements about the relative performance of the different strategies and so we now consider introducing some model assumptions in order to approximate the different components of the MSE expressions, in some specific settings. We first consider the additive model:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk},$$

where y_{ijk} is the k-th observation in the ij-th cell, α_i and β_j are fixed effects and ϵ_{ijk} are independent errors with zero mean and common variance σ^2. Then $E_m(S_{ij}^2) = \sigma^2$ and $E_m(\bar{Y}_{ij}\bar{Y}_{i'j'}) \doteq (\mu + \alpha_i + \beta_j)(\mu + \alpha_{i'} + \beta_{j'})$. Thus the model-expected design-variance is given by replacing S_{ij}^2 by σ^2 and $\bar{Y}_{ij}$ by $\mu + \alpha_i + \beta_j$ in the formulas for v_1 and v_2 for the various estimators. In this case, $v_2(\bar{y}_B) = 0$. This point was realized by Bryant *et al.* when comparing $\bar{y}_U$ and $\bar{y}_B$. The bias term will be zero in this case unless there was rounding on the margins, that is bias$(\bar{y}_B) = 0$ provided $n_{i\cdot} = nP_{i\cdot}$ and $n_{\cdot j} = nP_{\cdot j}$ as is the case in their example. This easily follows from (4.8) and

$$\sum_i (P_{ij} - P_{i\cdot}P_{\cdot j}) = \sum_j (P_{ij} - P_{i\cdot}P_{\cdot j}) = 0.$$

This was also shown by Bryant *et al.* p. 119 equation (47). Using (4.7), it is easily shown that $v_2(\bar{y}) = 0$ as well. This combined with the unbiasedness of $\bar{y}$ and the fact that $v_1(\bar{y}_B) = v_1(\bar{y}) = \sigma^2/n$ in this case implies that for this situation MSE$(\bar{y}_B)$ = MSE$(\bar{y})$, that is the proposed procedure has the same MSE as the procedure of Bryant *et al.* using the biased estimator. We demonstrate in the sequel that even when this additive model is applicable ($\gamma = 0$ below), $v_2(\bar{y}_U)$ may be large while $v_1(\bar{y}_U) > v_1(\bar{y})$.

To compare the estimators further, let us consider the situation of Example 1. The above derivations allow us to obtain the MSE's of the three estimators for this example provided we have the S_{ij}'s, the $\bar{Y}_{ij}$'s and can calculate the Cov$(n_{ij}, n_{i'j'})$ for the Bryant *et al.* method as well as for our approach. The covariances for the Bryant *et al.* method are given in their paper in terms of the P_{ij}'s, while the covariances for our approach can be obtained from the solution in Table 3. We will consider non-additive departures from the above model, namely

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma\alpha_i\beta_j + \epsilon_{ijk},$$

for various values of γ. For simplicity of presentation, let $\mu = 1$, $\alpha_i = i - 3$, $\beta_j = j - 2$ (note in fact that the MSE of each strategy is invariant to the choice of μ). Thus the model-expected design-variance is given by replacing S_{ij}^2 by 1 and $\bar{Y}_{ij}$ by $1 + (i-3) + (j-2) + \gamma(i-3)(j-2)$ in the formulas for v_1 and v_2 for the various estimators. Table 7 gives the resulting v_1, v_2, and MSE values for the three estimators (as well as the bias squared term for $\bar{y}_B$), for various values of γ. From Table 7, it can be seen that for an additive model, $\gamma = 0$, $\bar{y}_B$ and $\bar{y}$ perform equally well, while $\bar{y}_U$ is inferior. As the model becomes more non-additive, and $|\gamma|$ increases, the two estimators for the Bryant *et al.* strategy tend to perform similarly, both with MSE becoming increasingly greater than that of the proposed strategy. This pattern is primarily due to the v_2 component of the MSE of the three estimators. The bias term of $\bar{y}_B$ is of lesser importance, although it may be more important for larger n.

The greater increase in v_2 as $|\gamma|$ increases for the Bryant *et al.* design appears to reflect the greater variability of each n_{ij} for this design. It should be noted that it would have been possible to reduce this variability somewhat by applying a variant of the Bryant *et al.* method instead to Table 2, as was done for the proposed method, though one would need to derive adjusted G_{ij} weights for $\bar{y}_U$ and it would be difficult to handle the 0.0 cell entries in Table 2. However, even if this were accomplished, the $\tilde{n}_{ij}$ for this design may still take values other than just 0 and 1; for example n_{42} could take values 0, 1, or 2. This inflated n_c variability is inherent in the Bryant *et al.* method. For example, suppose $n_{1\cdot} = n_{\cdot 1} = 5$. Then using the Bryant *et al.* method, n_{11} can take values 0, 1, 2, 3, 4, or 5, while with the proposed method it can take only values $[nP_{11}]$ or $[nP_{11}] + 1$. If $nP_{11} < 1$, the technique used to go from Table 1 to Table 2 will not improve matters.

Table 7

Comparison of MSE for Three Estimators

	Bryant, Hartley, Jessen Design							Proposed Design		
	$\bar{y}_U$			$\bar{y}_B$				$\bar{y}$		
γ	v_1	v_2	MSE	v_1	v_2	Bias^2	MSE	v_1	v_2	MSE
0	.125	.105	.230	.100	.000	.000	.100	.100	.000	.100
±.5	.125	.063	.188	.100	.033	.002	.135	.100	.018	.118
±1	.125	.105	.230	.100	.131	.008	.239	.100	.071	.171
±2	.125	.440	.565	.100	.523	.032	.655	.100	.284	.384
±3	.125	1.111	1.236	.100	1.176	.073	1.349	.100	.638	.738

5. VARIANCE ESTIMATION

In this section, we will consider variance estimation for our proposed method. Using (4.1) and recalling constraint (2.1), it is clear that

$$\pi_{ck} = E[n_c(s)/N_c] = n/N.$$

The joint inclusion probability of two units k, k' in the same cell c is

$$\pi_{ck,ck'} = E[n_c(s)\{n_c(s) - 1\}/\{N_c(N_c - 1)\}].$$

Suppose $n_c(s) = I_c + \tilde{n}_c(s)$ when I_c is the fixed integer $[nP_c]$ and $\tilde{n}_c(s) = 0$ or 1.

If $nP_c \leq 1$ then $I_c = 0$ and $\pi_{ck,ck'} = 0$. Hence a necessary condition for unbiased variance estimation to be possible is that $nP_c > 1$ for all cells c. On the other hand if this condition holds then $n_c(s) \geq 1$ for all c and hence the probability of inclusion of any pair of units in different cells is also always positive. Hence this condition is necessary and sufficient for unbiased variance estimation to be possible.

When this condition holds we obtain

$$\pi_{ck,ck'} = I_c(I_c + 2r_c - 1)/[N_c(N_c - 1)] = A_c,$$

say, where $r_c = E[\tilde{n}_c(s)] = nP_c - I_c$.

The joint inclusion probability for pairs of units in different cells c and c' are

$$\begin{aligned}\pi_{ck,ck'} &= E[n_c(s)n_{c'}(s)/(N_cN_{c'})]\\ &= [I_cI_{c'} + r_{c'}I_c + r_cI_{c'} + r_{cc'}]/(N_cN_{c'}) = B_{cc'},\end{aligned} \tag{5.1}$$

say where $r_{cc'} = E[\tilde{n}_c(s)\tilde{n}_{c'}(s)]$.

Hence an unbiased estimator of $V(\bar{y}(s))$ of Sen-Yates-Grundy form may be constructed in the usual way.

In practice, however, we wish to consider situations where $nP_c \leq 1$ for some c. In this case one assumption we might make following Bryant *et al.* (1960, Sect. 7) in order to derive a variance estimator is that the population variance of Y is constant within each cell c, say S^2.

Let us first obtain the variance of $\bar{y}(s)$ in the general case

$$V(\bar{y}(s)) = \frac{1}{2n^2}\sum_c\sum_{k\neq k'}\sum\left(\frac{n^2}{N^2} - A_c\right)(y_{ck} - y_{ck'})^2 + \frac{1}{2n^2}\sum_{c\neq c'}\sum_{k,k'}\sum\left(\frac{n^2}{N^2} - B_{cc'}\right)(y_{ck} - y_{c'k'})^2.$$

Now providing $B_{cc'} > 0 \;\forall\; c, c'$ we may estimate the second term unbiasedly by

$$\frac{1}{2n^2}\sum_A\sum\sum_{k=1}^{n_c(s)}\sum_{k'=1}^{n_{c'}(s)}\left(\frac{\frac{n^2}{N^2} - B_{cc'}}{B_{cc'}}\right)(y_{ck} - y_{c'k'})^2,$$

where $A = \{c,c' : n_c(s) \geq 1, n_{c'}(s) \geq 1, c \neq c'\}$.

The first term can be written as

$$\frac{1}{2n^2}\sum_c\left(\frac{n^2}{N^2} - A_c\right)2N_c^2S^2.$$

For any c s.t. $n_c(s) \geq 2$

$$E\left(\sum_{k=1}^{n_c(s)}\sum_{\substack{k'=1\\k\neq k'}}^{n_c'(s)}\frac{(y_{ck} - y_{ck'})^2}{2n_c(s)\{n_c(s) - 1\}}\;\middle|\; n_c(s)\right) = S^2.$$

Thus provided at least one $n_c(s)$ is ≥ 2 an unbiased estimator of the first term is

$$\frac{1}{2n^2D}\sum_{\{c:n_c(s)\geq 2\}}\left(\frac{n^2}{N^2} - A_c\right)2N_c^2\sum_{k=1}^{n_c(s)}\sum_{k'=1}^{n_c'(s)}\frac{(y_{ck} - y_{ck'})^2}{2n_c(s)\{n_c(s) - 1\}}$$

where D = the number of cells, c, such that $n_c(s) \geq 2$.

The above requires $B_{cc'} > 0$. If

$$I_c = I_{c'} = 0,$$

by (5.1), we need

$$r_{cc'} = \sum \tilde{n}_c(s)\tilde{n}_{c'}(s)p(s) > 0, \tag{5.2}$$

which is linear in $p(s)$. The constraint (5.2) can be handled in linear programming if desired. There will be such a constraint for each pair c, c' s.t. $I_c = I_{c'} = 0$.

6. CONCLUDING REMARKS

We have proposed a linear programming approach to multi-way stratification, applying ideas of Rao and Nigam (1990, 1992). The approach is simple in conception and is very flexible in allowing for a range of different objectives via the loss function $w(s)$, as well as in permitting a variety of constraints such as that the joint inclusion probabilities of all cells be positive. The main practical constraint on the procedure is that it may rapidly become computationally expensive if not impossible as the number of cells in the multi-way classification increases. Some ideas on how to reduce the amount of computation have been considered. Further research on this question would be useful. For cases where the computational demands are prohibitive, the method of Causey *et al.* (1985) remains an alternative.

ACKNOWLEDGEMENTS

The authors wish to express their thanks to Wesley Yung for his computer programming assistance and J.N.K. Rao for helpful comments in the early stages of discussion. R.R. Sitter is supported by the Natural Sciences and Engineering Council of Canada. Our thanks are due to the referees and the Associate Editor for constructive suggestions and comments.

APPENDIX

Proof of (4.7) for Proposed Method

Note that

$$\mathrm{Cov}(n_{ij}(s), n_{i'j'}(s)) = \mathrm{E}(n_{ij}(s) n_{i'j'}(s)) - \mathrm{E}(n_{ij}(s))\mathrm{E}(n_{i'j'}(s)).$$

Equation (2.1) states that $\mathrm{E}(n_{ij}(s)) = nP_{ij}$. By definition

$$\mathrm{E}(n_{ij}(s) n_{i'j'}(s)) = \sum_s n_{ij}(s) n_{i'j'}(s) p(s).$$

Thus

$$\sum_j \mathrm{E}(n_{ij}(s))\mathrm{E}(n_{i'j'}(s)) = n^2 P_{i'j'} \sum_j P_{ij} = n^2 P_{i'j'} P_{i\cdot}, \tag{7.1}$$

and

$$\sum_j \mathrm{E}(n_{ij}(s) n_{i'j'}(s)) = \sum_j \sum_s n_{ij}(s) n_{i'j'}(s) p(s)$$
$$= \sum_s p(s) n_{i'j'}(s) \sum_j n_{ij}(s). \tag{7.2}$$

Assume that the solution to the linear optimization problem (2.2) equals zero, where $w(s)$ is given in (2.9). In this case, $\sum_j n_{ij}(s) = n_{i\cdot}(s) = nP_{i\cdot}$ and (7.2) implies

$$\sum_j \mathrm{E}(n_{ij}(s) n_{i'j'}(s)) = \sum_s p(s) n_{i'j'}(s) nP_{i\cdot}$$
$$= nP_{i\cdot} \sum_s n_{i'j'}(s) p(s)$$
$$= nP_{i\cdot}\mathrm{E}(n_{i'j'}(s)) = nP_{i\cdot} nP_{i'j'}. \tag{7.3}$$

Equations (7.1) and (7.3) together imply $\sum_j \mathrm{Cov}(n_{ij}(s), n_{i'j'}(s)) = 0$. It can be similarly shown that

$$\sum_i \mathrm{Cov}(n_{ij}(s), n_{i'j'}(s)) = \sum_{i'} \mathrm{Cov}(n_{ij}(s), n_{i'j'}(s))$$
$$= \sum_{j'} \mathrm{Cov}(n_{ij}(s), n_{i'j'}(s)) = 0.$$

REFERENCES

BRYANT, E.C., HARTLEY, H.O., and JESSEN, R.J. (1960). Design and estimation in two-way stratification. *Journal of the American Statistical Association*, 55, 105-124.

CAUSEY, B.D., COX, L.H., and ERNST, L.R. (1985). Applications of transportation theory to statistical problems. *Journal of the American Statistical Association*, 80, 903-909.

GOODMAN, R., and KISH, L. (1950). Controlled selection-a technique in probability sampling. *Journal of the American Statistical Association*, 45, 350-372.

JESSEN, R.J. (1970). Probability sampling with marginal constraints. *Journal of the American Statistical Association*, 65, 776-795.

RAO, J.N.K., and NIGAM, A.K. (1990). Optimal controlled sampling design. *Biometrika*, 77, 807-814.

RAO, J.N.K., and NIGAM, A.K. (1992). 'Optimal' controlled sampling: a unified approach. *International Statistical Review*, 60, 89-98.

WATERTON, J.J. (1983). An exercise in controlled selection. *Applied Statistics*, 32, 150-164.