Domain estimation under design-based models

Viviana B. Lencina
Departamento de Investigación, FM
Universidad Nacional de Tucumán, Argentina

Julio M. Singer and Heleno Bolfarine
Departamento de Estatística, IME
Universidade de São Paulo, Brazil

Luz Mery González
Departamento de Estadística
Universidad Nacional de Colombia, Colombia

Abstract

We obtain an optimal estimator for a domain (small area) total using a linear least-squares prediction approach under a design-based model. The optimal estimator is the same as that obtained under a super-population model approach, and reproduces the classical form of the synthetic estimator in an extreme case. Using the concept of M-optimality, we generalize a well known theorem (Royall, 1976) originally considered for super-population models and extend the results to include the estimation of the vector of the different domain totals.

Keywords: small area estimation, design based models, optimal estimation

1 Introduction

It is often the case that domains (many times referred to as small areas) of interest are only specified after a survey has been designed and carried out. In such cases, the statistician′s dilemma is to produce accurate estimates without being given the resources to collect the necessary data, as highlighted by Holt et al. (1979). A reasonable solution is to employ data from other sources to improve the estimates. This is the problem attacked here. In particular, we seek accurate estimators of domain parameters of interest (means, totals, quantiles, etc.) based on very small samples taken therefrom.

Marker (1999) describes existing domain estimators in a stratified sampling setup and organizes them from a general linear regression perspective. He includes a derivation of the conditions under which it is possible to view synthetic estimation as a form of regression. He also observes how the best linear unbiased predictor (BLUP) for the average of a domain is obtained using a super-population model approach (Bolfarine and Zacks, 1992) with a one-way Analysis of Variance (ANOVA) model that assumes, as in synthetic estimation, that all individual units belonging to the same stratum have the same expected value, regardless of the domain where they come from. The BLUP coincides with the synthetic estimator when there are no sampled data from the domain of interest.

In this paper we study the problem of estimating domain totals in situations where the population, and consequently, each domain, is divided into strata and inference rests on a design-based model. We assume that the sampling scheme corresponds to stratified simple random sampling. We use this approach instead of the one-way ANOVA super-population model discussed in Holt et al. (1979).

In Section 2 we generalize Theorem 2.1 of Royall (1976) to obtain the best linear unbiased estimator (BLUE) for any linear combination of the individual parameters of the finite population, as well as for vector valued parameters of interest. In this setup, estimation is reduced to a prediction problem under a model induced by the sampling scheme. In Section 3, we describe the notation and derive the BLUE for a specific domain total as well as for the vector of domain totals. Some concluding comments are presented in Section 4.

2 Best linear unbiased estimation in finite populations

Consider a vector $\mathbf{y}$ of N known constants, each of which is associated to a labeled unit in a finite population. Consider also a vector of random variables $\mathbf{Y}$ linked to $\mathbf{y}$ through a probability model. For example, $\mathbf{Y}$ may denote a super-population from which $\mathbf{y}$ is a realization. Alternatively, the probability model may be induced by the sampling scheme and its parameters are explicitly associated to the finite population vector in such way that if we observe $\mathbf{Y}$ completely we will know all the individual parameters in $\mathbf{y}$.

Along the lines of Stanek et al. (2004), we focus our interest in a $p \times 1$ vector of parameters of the form $\boldsymbol{\theta} = \mathbf{G}_*^t \mathbf{y}$, where $\mathbf{G}_*$ is a $N \times p$ matrix of known constants. There are situations in which these parameters may also be written as $\boldsymbol{\theta} = \mathbf{G}^t \mathbf{Y}$, as when the distribution of $\mathbf{Y}$ corresponds to the typical random permutation model and our target parameter is the finite population total, so that $\mathbf{G}_* = \mathbf{G} = \mathbf{1}_N$, where $\mathbf{1}_a$ denotes a vector of dimension a, with all elements equal to 1. On the other hand, if our target is the individual parameter associated to a specific unit in the finite population, for example, unit 1, it follows that $\mathbf{G}_* = \mathbf{e}_1$, with $\mathbf{e}_1$ denoting an N-dimensional vector with value one in the first position and zero in the remaining positions. Then $\boldsymbol{\theta}$ may not be written as $\mathbf{G}^t \mathbf{Y}$ because the random permutation model does not identify the units or labels in the finite population. Also, we may note that there are situations in which some characteristic of interest could be written as $\mathbf{G}^t\mathbf{Y}$ but not as $\mathbf{G}_*^t\mathbf{y}$. This happens, for example, when our interest is to predict the random variable that will appear in the i^{th} position in a permutation, i.e. when $\mathbf{G} = \mathbf{e}_i$.

In this paper we consider finite population parameters that may be written as $\boldsymbol{\theta} = \mathbf{G}^t\mathbf{Y}$. Let $\mathbf{Y}_S$ denote the portion of $\mathbf{Y}$ that is observed after the sample is selected, and $\mathbf{Y}_R$ denote the remainder. We consider probability models for which there exists a permutation matrix $\mathbf{K} = [\mathbf{K}_S^t\, \mathbf{K}_R^t]^t$ such that

$$\begin{bmatrix} \mathbf{Y}_S \\ \mathbf{Y}_R \end{bmatrix} = \begin{bmatrix} \mathbf{K}_S \\ \mathbf{K}_R \end{bmatrix} \mathbf{Y}.$$

In this context, the target parameter may be written explicitly as a linear function of $\mathbf{Y}_S$ and $\mathbf{Y}_R$, specifically,

$$\boldsymbol{\theta} = \mathbf{G}_S^t\mathbf{Y}_S + \mathbf{G}_R^t\mathbf{Y}_R, \tag{1}$$

where $\mathbf{G}_S^t = \mathbf{G}^t\mathbf{K}_S^t$ and $\mathbf{G}_R^t = \mathbf{G}^t\mathbf{K}_R^t$. Consequently, the problem of estimating $\boldsymbol{\theta}$ is equivalent to that of predicting $\mathbf{G}_R^t\mathbf{Y}_R$, as pointed by Royall (1976) for the case in which the parameter of interest is the finite population total.

Using the same model as in Royall (1976), we may write

$$\begin{bmatrix} \mathbf{Y}_S \\ \mathbf{Y}_R \end{bmatrix} = \begin{bmatrix} \mathbf{X}_S \\ \mathbf{X}_R \end{bmatrix} \boldsymbol{\beta} + \mathbf{E}, \tag{2}$$

with $\mathbf{E}(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$, $\mathbf{X} = [\mathbf{X}_S^t\, \mathbf{X}_R^t]^t$ and $\mathbf{V}(\mathbf{Y}) = \begin{bmatrix} \mathbf{V}_S & \mathbf{V}_{SR} \\ \mathbf{V}_{RS} & \mathbf{V}_R \end{bmatrix}$ is partitioned in accordance to $\mathbf{Y}$.

When the parameter of interest is scalar, it is common to search for the estimator with minimal variance among unbiased estimators in some class $\mathcal{C}_E$. For vector valued parameters, there are different ways of defining the optimal estimator. We adopt the concept of **M-optimality** in the class of unbiased estimators, i.e., we say that $\hat{\boldsymbol{\theta}}^* \in \mathcal{C}_E$ is optimum if and only if $\hat{\boldsymbol{\theta}}^*$

is unbiased for $\boldsymbol{\theta}$ and $\mathbf{V}(\mathbf{k}^t\hat{\boldsymbol{\theta}}^*) \leq \mathbf{V}(\mathbf{k}^t\hat{\boldsymbol{\theta}})$ for all $\mathbf{k} \in \mathbf{R}^p$ and for all unbiased $\hat{\boldsymbol{\theta}} \in \mathcal{C}_E$. This concept of M-optimality is based on the *more concentrated* concept discussed in Lehmann and Casella (1998, p. 347). The next theorem enables us to obtain such an optimal estimator.

Theorem 1 *Consider the setup described in (2), and assume that* $\mathbf{X}_S$ *has full column rank. Among the linear estimators* $\hat{\boldsymbol{\theta}} = \mathbf{A}\mathbf{Y}_S$ *satisfying* $\mathbf{E}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \mathbf{0}$*, the M-optimal estimator of* $\boldsymbol{\theta}$ *is*

$$\hat{\boldsymbol{\theta}}^* = \mathbf{G}_S^t\mathbf{Y}_S + \mathbf{G}_R^t[\mathbf{X}_R\hat{\boldsymbol{\beta}} + \mathbf{V}_{RS}\mathbf{V}_S^{-1}(\mathbf{Y}_S - \mathbf{X}_S\hat{\boldsymbol{\beta}})], \tag{3}$$

where the weighted least-squares estimator $\hat{\boldsymbol{\beta}} = (\mathbf{X}_S^t\mathbf{V}_S^{-1}\mathbf{X}_S)^{-1}\mathbf{X}_S^t\mathbf{V}_S^{-1}\mathbf{Y}_S$ *is the BLUE of* $\boldsymbol{\beta}$*. The corresponding prediction variance of* $\hat{\boldsymbol{\theta}}^*$ *is*

$$\begin{aligned}\mathbf{E}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta})^t = \; & \mathbf{G}_R^t(\mathbf{V}_R - \mathbf{V}_{RS}\mathbf{V}_S^{-1}\mathbf{V}_{SR})\mathbf{G}_R + \\ & +\mathbf{G}_R^t(\mathbf{X}_R - \mathbf{V}_{RS}\mathbf{V}_S^{-1}\mathbf{X}_S)\times \\ & \times(\mathbf{X}_S^t\mathbf{V}_S^{-1}\mathbf{X}_S)^{-1}(\mathbf{X}_R - \mathbf{V}_{RS}\mathbf{V}_S^{-1}\mathbf{X}_S)^t\mathbf{G}_R.\end{aligned}$$

The proof is presented in the Appendix.

The BLUE for a real valued linear combination of $\mathbf{Y}$, i.e., $\theta = \mathbf{g}^t\mathbf{Y} = \mathbf{g}_S^t\mathbf{Y}_S + \mathbf{g}_R^t\mathbf{Y}_R$, may be directly obtained from Theorem 1. Thus, when the parameter of interest is scalar, the estimator obtained from (3) coincides with the BLUE obtained in Appendix B of Stanek and Singer (2004). We also note that Theorem 2.1 of Royall (1976) is a special case of Theorem 1 above, for $\mathbf{G} = \mathbf{1}_N$ and $\mathbf{G}_R = \gamma = \mathbf{1}_{N-n}$, with n denoting the sample size. Finally, we observe that to estimate a vector valued linear combination of $\mathbf{y}$, each component is estimated optimally.

It is useful to mention at this point that since the theorem does not depend on the source of randomness, it may be used either in a super-population setup or under a design-based model, the only requirement being the specification of the expectation and the covariance matrix. If $\mathbf{X}_S$ is not a full column rank matrix but there is an unbiased estimator of $\boldsymbol{\theta}$, the theorem is still valid if we use a generalized inverse, $(\mathbf{X}_S^t\mathbf{V}_S^{-1}\mathbf{X}_S)^-$, instead of $(\mathbf{X}_S^t\mathbf{V}_S^{-1}\mathbf{X}_S)^{-1}$. Note that a necessary and sufficient condition for the existence of an unbiased estimator of $\boldsymbol{\theta}$ is that $\mathbf{X}_S^t\mathbf{X}_S^{t-}\mathbf{X}_R^t\mathbf{G}_R^t = \mathbf{X}_R^t\mathbf{G}_R^t$.

3 Notation and terminology for domain setup

We consider a finite population divided into J strata, and let $\mathcal{P}_j = \{1, ..., N_j\}$ $j = 1, \ldots, J$ denote the set of labels of the N_j (assumed known) units in each stratum. Associated with each unit in stratum j, there is a fixed value (parameter) y_{jk}, $k = 1, ..., N_j$. Let $\mathbf{y}_j = (y_{j1} \cdots y_{jN_j})^t$ denote the vector of such fixed values. The elements in each stratum are classified into I mutually exclusive and exhaustive domains labeled $i = 1, ..., I$; these labels completely classify the units into IJ cells, and we suppose, as in Holt et al. (1979), that the population sizes N_{ij} are known from previous censuses or from other sources of accurate information. Note that $N_j = \sum_{i=1}^{I} N_{ij}$.

Following Stanek et al. (2004), for all $j = 1, ..., J$ and $v = 1, ..., N_j$, we consider

$$Y_{jv} = \sum_{k=1}^{N_j} U_{vk}^{(j)} y_{jk},$$

where $U_{vk}^{(j)}$ is 1 if the unit k in stratum j is in position v after a random permutation of its elements and $U_{vk}^{(j)} = 0$, otherwise. Consequently, Y_{jv} represents what is observed in position v of stratum j after the random permutation. The vector of random variables $\mathbf{Y}_j = (Y_{j1} \cdots Y_{jN_j})^t$ can be written as $\mathbf{Y}_j = \mathbf{U}^{(j)}\mathbf{y}_j$ with

$$\mathbf{U}^{(j)} = \begin{bmatrix} U_{11}^{(j)} & U_{12}^{(j)} & \cdots & U_{1N_j}^{(j)} \\ U_{21}^{(j)} & U_{22}^{(j)} & \cdots & U_{2N_j}^{(j)} \\ \vdots & \vdots & \ddots & \vdots \\ U_{N_j1}^{(j)} & U_{N_j2}^{(j)} & \cdots & U_{N_jN_j}^{(j)} \end{bmatrix}.$$

Given the random structure of $\mathbf{U}^{(j)}$, the corresponding expected value and variance are respectively given by

$$\mathbf{E}(\mathbf{Y}_j) = \beta_j \mathbf{1}_{N_j} \text{ and } \mathbf{V}(\mathbf{Y}_j) = \sigma_j^2 \mathbf{P}_{N_j}, \tag{4}$$

where $\mathbf{P}_{N_j} = \mathbf{I}_{N_j} - N_j^{-1}\mathbf{J}_{N_j}$ and $\mathbf{J}_{N_j} = \mathbf{1}_{N_j}\mathbf{1}_{N_j}^t$. The term $\beta_j = N_j^{-1}\mathbf{1}_{N_j}^t\mathbf{y}_j$ is the mean and $[(N_j-1)/N_j]\sigma_j^2 = N_j^{-1}\mathbf{y}_j^t\mathbf{P}_{N_j}\mathbf{y}_j$ is the corresponding finite population variance.

The vector of random variables can be compactly written as $\mathbf{Y} = (\mathbf{Y}_1^t \cdots \mathbf{Y}_J^t)^t$, and from (4), its expected value and covariance matrix may be, respectively, re-expressed as

$$\mathbf{E}(\mathbf{Y}) = \left[\bigoplus_{j=1}^{J} \mathbf{1}_{N_j}\right] \boldsymbol{\beta} \text{ and } \mathbf{V}(\mathbf{Y}) = \bigoplus_{j=1}^{J} \sigma_j^2 \mathbf{P}_{N_j},$$

with $\bigoplus_{j=1}^{N} \mathbf{A}_j$ denoting a block diagonal matrix with blocks given by $\mathbf{A}_j$ (Searle, 1982).

Consider simple random samples of size n_j, $j = 1, ..., J$, without replacement obtained independently in each stratum. Under the random permutation setup, this is equivalent to observing the first n_j elements of each $\mathbf{Y}_j$. Let n_{ij} denote the number of elements belonging to domain i observed in the sample of stratum j. For the time being, assume that all $n_{ij} > 0$ and $(N_{ij} - n_{ij}) > 0$. Since the selection of elements in each stratum is independent of the domain to which they belong and since in a random permutation the elements are interchangeable, we may suppose that the first n_{1j} elements in $\mathbf{Y}_j$ are from domain 1, the second n_{2j} elements are from domain 2, and so forth, up to the last n_{Ij} elements of the observed part of $\mathbf{Y}_j$, which are assumed to come from domain I. On the other hand, we assume the same structure for the $N_{ij} - n_{ij}$ non-observed elements of each domain.

Now we consider the problem of estimating certain domain characteristics, conditionally on the observation of $n_{ij}, i = 1, \ldots, I, j = 1, \ldots, J$. Initially, let the interest be focused on the total response for domain i, θ_i; next we consider the vector of all domain totals, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_I)^t$. In both cases, the parameter of interest may be written as a linear combination of the random vector $\mathbf{Y}$, i.e., $\boldsymbol{\theta} = \mathbf{G}_S^t \mathbf{Y}_S + \mathbf{G}_R^t \mathbf{Y}_R$. For the total response of domain i,

$$\mathbf{G}_S^t = (\mathbf{1}_J^t \otimes \mathbf{e}_i^t) \bigoplus_{j=1}^{J} \bigoplus_{i=1}^{I} \mathbf{1}_{n_{ij}}^t \text{ and } \mathbf{G}_R^t = (\mathbf{1}_J^t \otimes \mathbf{e}_i^t) \bigoplus_{j=1}^{J} \bigoplus_{i=1}^{I} \mathbf{1}_{N_{ij}-n_{ij}}^t, \qquad (5)$$

and for the vector of domain totals,

$$\mathbf{G}_S^t = (\mathbf{1}_J^t \otimes \mathbf{I}_I) \bigoplus_{j=1}^{J} \bigoplus_{i=1}^{I} \mathbf{1}_{n_{ij}}^t \text{ and } \mathbf{G}_R^t = (\mathbf{1}_J^t \otimes \mathbf{I}_I) \bigoplus_{j=1}^{J} \bigoplus_{i=1}^{I} \mathbf{1}_{N_{ij}-n_{ij}}^t. \qquad (6)$$

To estimate the parameters we use the techniques described in Section 2, under model (2) with

$$\mathbf{X}_S = \bigoplus_{j=1}^{J} \mathbf{1}_{n_j}, \quad \text{and} \quad \mathbf{X}_R = \bigoplus_{j=1}^{J} \mathbf{1}_{N_j - n_j}.$$

Then,

$$\mathbf{V}_S = \bigoplus_{j=1}^{J} \sigma_j^2 (\mathbf{I}_{n_j} - \frac{1}{N_j} \mathbf{J}_{n_j}), \mathbf{V}_{SR} = \bigoplus_{j=1}^{J} \frac{-1}{N_j} \sigma_j^2 \mathbf{J}_{n_j \times (N_j - n_j)}, \text{ and}$$

$$\mathbf{V}_R = \bigoplus_{j=1}^{J} \sigma_j^2 (\mathbf{I}_{N_j - n_j} - \frac{1}{N_j} \mathbf{J}_{N_j - n_j}).$$

The estimator of the total of domain i may be obtained by applying Theorem 1 and is given by

$$\hat{\theta}_i^* = \left[(\mathbf{1}_J^t \otimes \mathbf{e}_i^t) \bigoplus_{j=1}^{J} \bigoplus_{i=1}^{I} \mathbf{1}_{n_{ij}}^t \right] \mathbf{Y}_S + \left[(\mathbf{1}_J^t \otimes \mathbf{e}_i^t) \bigoplus_{j=1}^{J} \frac{1}{n_j} (\mathbf{N}_{\star j} - \mathbf{n}_{\star j}) \mathbf{1}_{n_j}^t \right] \mathbf{Y}_S, \tag{7}$$

with $\mathbf{N}_{\star j} = (N_{1j}, N_{2j}, \cdots, N_{Ij})^t$ and $\mathbf{n}_{\star j} = (n_{1j}, n_{2j}, \cdots, n_{Ij})^t$. The mean squared error of this predictive estimator is

$$MSE(\hat{\theta}_i^*) = \sum_{j=1}^{J} \sigma_j^2 \frac{(N_{ij} - n_{ij})}{n_j} (N_{ij} - n_{ij} + n_j). \tag{8}$$

If $\boldsymbol{\theta}$ is the vector of totals for the domains, using (6) and (1), an application of Theorem 1 leads to the estimator

$$\hat{\boldsymbol{\theta}}^* = \left[(\mathbf{1}_J^t \otimes \mathbf{I}_I) \bigoplus_{j=1}^{J} \bigoplus_{i=1}^{I} \mathbf{1}_{n_{ij}}^t \right] \mathbf{Y}_S + \left[(\mathbf{1}_J^t \otimes \mathbf{I}_I) \bigoplus_{j=1}^{J} \frac{1}{n_j} (\mathbf{N}_{\star j} - \mathbf{n}_{\star j}) \mathbf{1}_{n_j}^t \right] \mathbf{Y}_S.$$

Under this approach, the total response for each domain is optimally estimated. The corresponding matrix of mean squared errors and cross products is

$$\begin{aligned} MSE(\hat{\boldsymbol{\theta}}^*) &= \mathbf{E}(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}^* - \boldsymbol{\theta})^t \\ &= \sum_{j=1}^{J} \sigma_j^2 \left[\mathbf{D}_{\mathbf{N}_{\star j} - \mathbf{n}_{\star j}} + \frac{1}{n_j} (\mathbf{N}_{\star j} - \mathbf{n}_{\star j})(\mathbf{N}_{\star j} - \mathbf{n}_{\star j})^t \right], \end{aligned} \tag{9}$$

with $\mathbf{D}_{\mathbf{x}}$ denoting a diagonal matrix with the elements of the vector $\mathbf{x}$ along the main diagonal.

When $n_{ij} = 0$ or $(N_{ij} - n_{ij}) = 0$ for some strata j and domain i, expressions (5) and (6) must be modified. First, let $\tilde{\mathbf{n}}_{\star j}$ contain only the terms with $n_{ij} > 0$ and $\mathbf{a}_{iS}$ be a vector obtained from $\tilde{\mathbf{n}}_{\star} = \begin{pmatrix} \tilde{\mathbf{n}}_{\star 1}^t & \dots & \tilde{\mathbf{n}}_{\star J}^t \end{pmatrix}^t$ by replacing the terms n_{ij} by 1 and $n_{i'j}, i' \neq i$, by 0 for all j. Also, let $\tilde{\mathbf{N}}_{\star j} - \tilde{\mathbf{n}}_{\star j}$ and $\mathbf{a}_{iR}$ be constructed similarly for the cells corresponding to $(N_{ij} - n_{ij}) > 0$.

Then for the total of domain i, (5) must be replaced by

$$\mathbf{G}_S^t = \mathbf{a}_{iS}^t \left(\bigoplus_{j=1}^{J} \bigoplus_{i \in \tilde{S}_j} \mathbf{1}_{n_{ij}}^t \right) \quad \text{and} \quad \mathbf{G}_R^t = \mathbf{a}_{iR}^t \left(\bigoplus_{j=1}^{J} \bigoplus_{i \in \tilde{R}_j} \mathbf{1}_{N_{ij}-n_{ij}}^t \right) \tag{10}$$

where $\tilde{S}_j = \{i : n_{ij} > 0\}$ and $\tilde{R}_j = \{i : (N_{ij} - n_{ij}) > 0\}$ and for the vector of totals, (6) must be replaced by

$$\mathbf{G}_S^t = \begin{pmatrix} \mathbf{a}_{1S}^t \\ \vdots \\ \mathbf{a}_{IS}^t \end{pmatrix} \left(\bigoplus_{j=1}^{J} \bigoplus_{i \in \tilde{S}_j} \mathbf{1}_{n_{ij}}^t \right) \quad \text{and} \quad \mathbf{G}_R^t = \begin{pmatrix} \mathbf{a}_{1R}^t \\ \vdots \\ \mathbf{a}_{IR}^t \end{pmatrix} \left(\bigoplus_{j=1}^{J} \bigoplus_{i \in \tilde{R}_j} \mathbf{1}_{N_{ij}-n_{ij}}^t \right). \tag{11}$$

As a result, the required estimators are

$$\hat{\theta}_i^* = \left[\mathbf{a}_{iS}^t \left(\bigoplus_{j=1}^{J} \bigoplus_{i \in \tilde{S}_j} \mathbf{1}_{n_{ij}}^t \right) \right] \mathbf{Y}_S + \left[\mathbf{a}_{iR}^t \left(\bigoplus_{j=1}^{J} \frac{1}{n_j} \left(\tilde{\mathbf{N}}_{\star j} - \tilde{\mathbf{n}}_{\star j} \right) \mathbf{1}_{n_j}^t \right) \right] \mathbf{Y}_S \tag{12}$$

and

$$\hat{\boldsymbol{\theta}}^* = \left[\begin{pmatrix} \mathbf{a}_{1S}^t \\ \vdots \\ \mathbf{a}_{IS}^t \end{pmatrix} \left(\bigoplus_{j=1}^{J} \bigoplus_{i \in \tilde{S}_j} \mathbf{1}_{n_{ij}}^t \right) \right] \mathbf{Y}_S + \left[\begin{pmatrix} \mathbf{a}_{1R}^t \\ \vdots \\ \mathbf{a}_{IR}^t \end{pmatrix} \left(\bigoplus_{j=1}^{J} \frac{1}{n_j} \left(\tilde{\mathbf{N}}_{\star j} - \tilde{\mathbf{n}}_{\star j} \right) \mathbf{1}_{n_j}^t \right) \right] \mathbf{Y}_S, \tag{13}$$

their corresponding mean square errors being given by (8) and (9), respectively. For illustrative purposes, an example with $J = 2$ and $I = 3$ is detailed in the Appendix.

4 Comments

We have shown that the BLUP obtained under the one-way ANOVA model may be reproduced under a design-based model, without any restrictions besides those induced by the sampling scheme. Since it is easier to attach face value to such a model than to some other super-population models, we feel that the estimators of the totals are better justified under the former approach. Additionally, the design-based model facilitates the definition of the parameters under investigation.

Previously, only the super-population model approach (Holt et al. 1979, Royall, 1976, Bolfarine and Zacks, 1992) was considered in the literature. We attacked the problem from a purely design-based approach and obtained optimal estimators that provide a simple rationale to previously used domain estimators. The proposed estimator is intuitive in the sense that if the population has a single stratum ($J = 1$), the estimator of the total for domain i reduces to $n_i\bar{y}_i + (N_i - n_i)\bar{y}$, that is, it estimates the response for the units of domain i that are not in the sample using the observed overall mean ($\bar{y}$) which encompasses information for all the other sampled domains (thus, "borrowing information" from other domains). Moreover, the variance of the estimator depends only on the finite population variances (σ_j^2) and not on artificial variances imposed by the super-population model.

It is easy to show that $\hat{\theta}_i^*$ coincides with

$$\hat{T}_i^* = \sum_{j=1}^{J} n_{ij}(\bar{y}_{ij\bullet} - \bar{y}_{\bullet j\bullet}) + \sum_{j=1}^{J} N_{ij}\bar{y}_{\bullet j\bullet},$$

as obtained by Holt et al. (1979) under a one-way ANOVA superpopulation model, where $\bar{y}_{ij\bullet}$ denotes the average response among all the elements in the sample that belong to domain i and stratum j and $\bar{y}_{\bullet j\bullet}$ denotes the response average of all the elements in the sample that belong to stratum j.

Furthermore, when there are no sampled data from the domain of interest, it coincides with the synthetic estimator, which for the total θ_i of domain i is

$$\hat{T}_i = \sum_{j=1}^{J} N_{ij}\overline{y}_{\bullet j \bullet}.$$

The accuracy of this estimator depends on the similarity between each stratum and the domain and also on the accuracy of the weights. In situations for which N_{ij} is much greater than n_{ij}, the difference between both estimators is immaterial.

Acknowledgements

We are grateful to Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), FINEP (PRONEX) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil, Consejo de Investigaciones de la Universidad Nacional de Tucumán (CIUNT), Argentina, and to the National Institutes of Health (NIH-PHS-R01-HD36848), United States, for financial support.

Appendix

Proof of Theorem 1: Let $\mathbf{k} \in \mathbf{R}^p$, so that $\theta = \mathbf{k}^t\boldsymbol{\theta} = \mathbf{k}^t\mathbf{G}_S^t\mathbf{Y}_S + \mathbf{k}^t\mathbf{G}_R^t\mathbf{Y}_R$ is a real valued parameter. As shown by Stanek and Singer (2002, Appendix B), the BLUE of θ is

$$\hat{\theta}^* = \mathbf{k}^t\mathbf{G}_S^t\mathbf{Y}_S + \mathbf{k}^t\mathbf{G}_R^t[\mathbf{X}_R\hat{\boldsymbol{\beta}} + \mathbf{V}_{RS}\mathbf{V}_S^{-1}(\mathbf{Y}_S - \mathbf{X}_S\hat{\boldsymbol{\beta}})] = \mathbf{k}^t\hat{\boldsymbol{\theta}}^*.$$

Accordingly, for every linear unbiased estimator of θ, i.e., $\hat{\theta} = \mathbf{a}^t\mathbf{Y}_S$ with $\mathbf{a}$ denoting a vector of constants, it follows that

$$\mathbf{V}(\hat{\theta}) = \mathbf{V}(\mathbf{k}^t\hat{\boldsymbol{\theta}}^*) \leq \mathbf{V}(\mathbf{a}^t\mathbf{Y}_S).$$

For all linear unbiased estimators of $\boldsymbol{\theta}$, i.e., $\hat{\boldsymbol{\theta}} = \mathbf{A}\mathbf{Y}_S$ with $\mathbf{A}$ denoting a matrix of constants, it follows that $\mathbf{k}^t\mathbf{A}\mathbf{Y}_S$ is a linear unbiased estimator of θ, so that

$$\mathbf{V}(\mathbf{k}^t\hat{\boldsymbol{\theta}}^*) \leq \mathbf{V}(\mathbf{k}^t\hat{\boldsymbol{\theta}}) \text{ for all } \hat{\boldsymbol{\theta}} = \mathbf{A}\mathbf{Y}_S \text{ and for all } \mathbf{k} \in \mathbf{R}^p.$$

Consequently, $\hat{\boldsymbol{\theta}}^*$ is the M-optimal estimator of $\boldsymbol{\theta}$.

The corresponding predictor variance of $\hat{\boldsymbol{\theta}}^*$ follows easily after some matrix operations.

Example: In Tables 1 and 2, we illustrate a finite population with $J = 2$, $I = 3$, and the corresponding sample sizes.

Table 1: Finite population.

Strata (j)	Domain (i)	N_{ij}	
1	1	2	y_{11} y_{14}
	2	1	y_{13}
	3	2	y_{12} y_{15}
2	1	0	
	2	3	y_{22} y_{23} y_{25}
	3	3	y_{21} y_{24} y_{26}

In this case, $\boldsymbol{y} = \begin{pmatrix} y_{11} & \ldots & y_{15} & y_{21} & \ldots & y_{26} \end{pmatrix}^t$ and $\boldsymbol{Y} = \begin{pmatrix} \boldsymbol{Y}_1^t & \boldsymbol{Y}_2^t \end{pmatrix}^t = \begin{pmatrix} Y_{11} & \ldots & Y_{15} & Y_{21} & \ldots & Y_{26} \end{pmatrix}^t$.

To estimate the total of domain i, (10) and (12) are specified by taking

$$\tilde{S}_1 = \{1, 2\}, \tilde{S}_2 = \{2\}, \tilde{R}_1 = \{1, 3\}, \tilde{R}_2 = \{2, 3\},$$

Table 2: Sample sizes.

Strata (j)	Domain (i)	n_{ij}	$N_{ij}-n_{ij}$
1	1	1	1
	2	1	0
	3	0	2
2	1	0	0
	2	2	1
	3	0	3
Total		4	7

$$\bigoplus_{j=1}^{2}\bigoplus_{i\in\tilde{S}_j}\mathbf{1}_{n_{ij}}^t = \begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 1\end{pmatrix}, \bigoplus_{j=1}^{2}\bigoplus_{i\in\tilde{R}_j}\mathbf{1}_{N_{ij}-n_{ij}}^t = \begin{pmatrix}1 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 1 & 1 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 1 & 1 & 1\end{pmatrix},$$

$$\left(\tilde{\mathbf{N}}_{\star 1} - \tilde{\mathbf{n}}_{\star 1}\right) = \begin{pmatrix}1\\2\end{pmatrix}, \quad \left(\tilde{\mathbf{N}}_{\star 2} - \tilde{\mathbf{n}}_{\star 2}\right) = \begin{pmatrix}1\\3\end{pmatrix},$$

$$\mathbf{Y}_S = \begin{pmatrix}Y_{11} & Y_{12} & Y_{21} & Y_{22}\end{pmatrix}^t, \quad \mathbf{Y}_R = \begin{pmatrix}Y_{13} & Y_{14} & Y_{15} & Y_{23} & Y_{24} & Y_{25} & Y_{26}\end{pmatrix}^t,$$

$$\mathbf{a}_{1S} = \begin{pmatrix}1 & 0 & 0\end{pmatrix}^t, \quad \mathbf{a}_{1R} = \begin{pmatrix}1 & 0 & 0 & 0\end{pmatrix}^t,$$

$$\mathbf{a}_{2S} = \begin{pmatrix}0 & 1 & 1\end{pmatrix}^t, \quad \mathbf{a}_{2R} = \begin{pmatrix}0 & 0 & 1 & 0\end{pmatrix}^t \text{ and}$$

$$\mathbf{a}_{3S} = \begin{pmatrix}0 & 0 & 0\end{pmatrix}^t, \quad \mathbf{a}_{3R} = \begin{pmatrix}0 & 1 & 0 & 1\end{pmatrix}^t.$$

This results in the following estimators and corresponding MSE

$$\hat{\theta}_1^* = Y_{11} + \frac{1}{2}\left(Y_{11} + Y_{12}\right), \quad MSE\left(\hat{\theta}_1^*\right) = \frac{3}{2}\sigma_1^2,$$

$$\hat{\theta}_2^* = Y_{12} + \frac{3}{2}\left(Y_{21} + Y_{22}\right), \quad MSE\left(\hat{\theta}_2^*\right) = \frac{3}{2}\sigma_2^2, \text{ and}$$

$$\hat{\theta}_3^* = \left(Y_{11} + Y_{12}\right) + \frac{3}{2}\left(Y_{21} + Y_{22}\right), \quad MSE\left(\hat{\theta}_3^*\right) = 4\sigma_1^2 + \frac{15}{2}\sigma_2^2.$$

For the vector of totals, from (9) with $(\boldsymbol{N}_{\star 1} - \boldsymbol{n}_{\star 1}) = \begin{pmatrix} 1 & 0 & 2 \end{pmatrix}^t$ and $(\boldsymbol{N}_{\star 2} - \boldsymbol{n}_{\star 2}) = \begin{pmatrix} 0 & 1 & 3 \end{pmatrix}^t$, we obtain

$$MSE\left(\hat{\boldsymbol{\theta}}^*\right) = \begin{pmatrix} \frac{3}{2}\sigma_1^2 & 0 & \sigma_1^2 \\ 0 & \frac{3}{2}\sigma_2^2 & \frac{3}{2}\sigma_2^2 \\ \sigma_1^2 & \frac{3}{2}\sigma_2^2 & 4\sigma_1^2 + \frac{15}{2}\sigma_2^2 \end{pmatrix}.$$

References

Bolfarine, H. and Zacks, S. (1992) *Predition theory for finite population quantities*. Springer, N.Y.

Holt, D., Smith, T.F.M. and Tomberlin, T.J. (1979). A model based approach to estimation of small subgroups of a population. *Journal of the American Statistical association*, **74**, 405-410.

Lehmann, E.L. and Casella, G. (1998). *Theory of point estimation.* Springer, N.Y.

Marker, D.A. (1999). Organization of Small Areas Estimators Using a Generalized Linear Regression Framework. *Journal of Official Statistics*, **15** (1), 1-24.

ROYALL, R. (1976). The linear least squares prediction approach to two stage sampling. *Journal of the American Statistical Association*, **71**, 657-664.

SEARLE, S. (1982). *Matrix algebra useful for statistics*. Wiley, N.Y.

STANEK, E.J. III, SINGER, J.M. AND LENCINA, V.B. (2004). A unified approach to estimation and prediction under simple random sampling. *Journal of Statistical Planning and Inference*, **121**, 325-338.

STANEK, E.J. III AND SINGER, J.M. (2004). Predicting random effects from finite population clustered samples with response error. *Journal of the American Statistical Association*, **99**, 1119-1130.