Metrika, Volume 31, 1984, page 319–322.

On Brewer's Class of Robust Sampling Designs for Large-Scale Surveys

By *T.J. Rao*, Blacksburg[1])

Summary: In this note we observe that Brewer's (1979) result concerning the asymptotically design unbiased strategy which has minimum expected variance under a super population model can be established in a more general setting.

1. Introduction

Consider the problem of estimating the finite population total $Y = \sum_{i=1}^{N} Y_i$ of a characteristic y, taking values Y_i on the units U_i, $i = 1, 2, \ldots, N$. Suppose that information on an auxiliary characteristic x related to y is available on all the units of the population taking values X_i on U_i, $i = 1, 2, \ldots, N$. Given a sample of n distinct units, the estimation problem can then be considered as one of estimating $Y - \sum_{i \in s} Y_i$, the total of the $(N - n)$ unobserved units. For this, the following super population model is used where we assume

$$Y_i = \beta X_i + e_i,$$

$$E(e_i) = 0, E(e_i e_j) = \sigma_i^2 \text{ if } j = i \qquad (1.1)$$
$$= 0, \text{ otherwise,}$$

and a class of predictors for Y is given by

$$y^* = \sum_{i \in s} Y_i + \hat{\beta} \sum_{i \notin s} X_i. \qquad (1.2)$$

Brewer [1979] has suggested the use of the model unbiased estimator

$$\hat{\beta}_B = \sum_{i \in s} W_i Y_i / \sum_{i \in s} W_i X_i$$

---

[1]) Dr. *T.J. Rao*, Virginia Polytechnic Institute and State University, Blacksburg, VA, U.S.A. On leave from Indian Statistical Institute, Calcutta. This work was done while the author was visiting Iowa State University, Ames.

0026-1335/84/030319-322$2.50 © *1984 Physica-Verlag, Vienna.*

where W_i's are arbitrary, instead of the best linear unbiased (BLU) estimator of β given by

$$\hat{\beta}_{\mathrm{BLU}} = \sum_{i \in s} Y_i X_i \sigma_i^{-2} / \sum_{i \in s} X_i^2 \sigma_i^{-2}.$$

He then requires that y^* be asymptotically unbiased over repeated sampling. We adopt the following formulation used by *Brewer* [1979] in his asymptotic analysis. The original population of N units is reproduced $(k-1)$ times, yielding k identical populations of N units each. From each of the k populations, a sample is selected using the same π_i for each one. The k populations are then aggregated to an overall population of size $N_k = Nk$ whose population total for the y-characteristic is $Y_k = kY$ and the k samples are aggregated to an overall sample of $n_k = nk$ units. The estimator y_k^* of the population total Y_k is now obtained from (1.1) using $\hat{\beta}_B$. k is then allowed to tend to infinity. Using these assumptions of the asymptotic analysis, *Brewer* [1979] then minimizes the asymptotic value of the expected variance of y_k^* / N_k under the model (1.1). This leads to the choice of optimum weights given by $\alpha(\pi_i^{-1} - 1)$, where $\pi_i \propto \sigma_i$ and α is a constant. In this paper, we generalize Brewer's estimator and give a slightly modified proof, following exactly the same spirit of the asymptotic analysis of Brewer's.

2. Main Results

Write $Y = \sum_{i \in s} Y_i + \beta (X - \sum_{i \in s} X_i)$

where

$$\beta = (Y - \sum_{i \in s} Y_i) / (X - \sum_{i \in s} X_i) \tag{2.1}$$

and consider the problem of estimating β. Following *Godambe* [1955], let

$$\hat{\beta}_G = \sum_{i \in s} (\beta_{si} - 1) Y_i / \sum_{i \in s} (\beta_{si} - 1) X_i. \tag{2.2}$$

Here β can be regarded as a weighted average of the unobserved ratios Y_i / X_i, weights being the sizes of the corresponding units. It is then natural [cf. *Basu*] to estimate β by some sort of a weighted average of the observed ratios Y_i / X_i, $i \in s$ and the weights used in (2.2) are $(\beta_{si} - 1) X_i$. Imitating the asymptotic analysis of Brewer, we how have

$$\lim_{k \to \infty} E_p (y_k^* / N_k) = N^{-1} [\sum_{i=1}^{N} \pi_i Y_i + (\sum_{i=1}^{N} Y_i (a_i - \pi_i) / \sum_{i=1}^{N} X_i (a_i - \pi_i)) \sum_{i=1}^{N} (1 - \pi_i) X \tag{2.3}$$

where $a_i = \sum_{s \in i} \beta_{si} p(s)$ and E_p denotes the expectation over the design. Thus y^* is asymptotically unbiased iff

$$\pi_i = (a_i - \alpha) / (1 - \alpha) \tag{2.4}$$

where

$$\alpha = \sum_{i=1}^{N} X_i (a_i - \pi_i) / \sum_{i=1}^{N} X_i (1 - \pi_i) \tag{2.5}$$

and $a_i \neq 1$. When $a_i = 1$, $\sum_{i \in s} \beta_{si} Y_i$ and $\sum_{i \in s} \beta_{si} X_i$ are unbiased estimators of Y and X respectively and the above condition of unbiasedness is automatically satisfied. Further, we have

$$V = \lim_{k \to \infty} E_p E \{(y_k^* - Y_k)^2 / N_k\}$$

$$= N^{-1} [((\sum_{i=1}^{N} x_i (1 - \pi_i))^2 / \{\sum_{i=1}^{N} X_i (a_i - \pi_i)\}^2) \sum_{i=1}^{N} \sigma_i^2 \sum_{s \in i} (\beta_{si} - 1)^2 p(s) +$$

$$+ \sum_{i=1}^{N} (1 - \pi_i) \sigma_i^2]. \tag{2.6}$$

Next, using the fact that

$$\sum_{s \in i} \beta_{si}^2 p(s) \geqslant a_i^2 / \pi_i \tag{2.7}$$

and substituting (2.4) in (2.6), we get

$$NV \geqslant \sum_{i=1}^{N} \sigma_i^2 \{(a_i^2 (1 - \alpha) / \alpha^2 (a_i - \alpha)) + ((2a_i \alpha - a_i - \alpha) / (1 - \alpha) \alpha^2) +$$

$$+ ((1 - a_i) / (1 - \alpha)) \tag{2.8}$$

$$= \sum_{i=1}^{N} \sigma_i^2 (1 - a_i) / (a_i - \alpha). \tag{2.9}$$

Equality in (2.7) and (2.8) occurs iff $\beta_{si} = a_i / \pi_i$. This condition is the same as $(\beta_{si} - 1) = (a_i - \pi_i) \pi_i^{-1} = \alpha (\pi_i^{-1} - 1)$. Notice that (2.9) is the Godambe Joshi lower bound $\sum_{i=1}^{N} \sigma_i^2 (\pi_i^{-1} - 1)$ and this can be minimized when π_i's are proportional to σ_i. Thus the minimum of the r.h.s. of (2.8) is attained when β_{si} are chosen to $1 + \alpha (n^{-1} \sigma_i^{-1} \sum_{i=1}^{N} \sigma_i - 1)$ and then the optimum estimator is given by

$$y^* = \sum_{i \in s} Y_i + (\sum_{i \in s} Y_i (\pi_i^{-1} - 1) / \sum_{i \in s} X_i (\pi_i^{-1} - 1)) \sum_{i \notin s} X_i$$

where $\pi_i = n\sigma_i / \sum_{i=1}^{N} \sigma_i$. It is interesting to note that α cancels off and does not enter the estimator.

Remark 2.1: The case when $a_i - 1$ easy to deal with and we shall omit the details here. (Notice that $a_i < 1$ in the above result, since $\pi_i < 1$.)

Remark 2.2: *Fuller/Isaki* [1980] have considered design consistent estimators which are not necessarily design unbiased and presented a strategy such that the predictor is, given the sample, best linear unbiased under the model. They have also given empirical examples to compare the ratio estimator

$$\hat{\bar{Y}}_R = (\sum_{i \in s} X_i / \pi_i)^{-1} (\sum_{i \in s} Y_i / \pi_i) \bar{X},$$

Brewer's estimator $\hat{\bar{Y}}_B = y^*$ considered above, and *Cassel/Särndal/Wretman* [1976] estimator

$$\hat{\bar{Y}}_{CSW} = \hat{\bar{Y}}_{HT} + \hat{\beta} (\bar{Y} - \hat{\bar{X}}_{HT})$$

where $\hat{\beta} = (\sum_{i \in s} X_i^2 \pi_i^{-2})^{-1} (\sum_{i \in s} Y_i X_i \pi_i^{-2})$ and the regression estimator. For further discussion we refer to *Fuller/Isaki* [1980].

References

Basu, D.: An Essay on the Logical Foundations of Survey Sampling, part one. In: Symposium on the Foundations of Statistical Inference. Ed. by V.P. Godambe, and D.A. Sprott. Toronto Montreal 1971, 203–242.

Brewer, K.R.W.: A Class of Robust Sampling Designs for Large-Scale Surveys. Journal of the American Statistical Association 74, 1979, 911–915.

Cassel, C.M., C.E. Särndal, and *J.H. Wretman*: Some Results on Generalized Difference Estimation and Generalized Regression Estimation for Finite Populations. Biometrika 63, 1976, 615–620.

Fuller, W.A., and *C.T. Isaki*: Survey Design under Super Population Models. To appear in Journal of the American Statistical Association, Iowa State University Research Report 1980.

Godambe, V.P.: A Unified Theory of Sampling from Finite Populations. Journal of the Royal Statistical Society Ser. B, 17, 1955, 269–278.

Received August 5, 1981
(revised version October 10, 1981)