BULLETIN of the MALAYSIAN MATHEMATICAL SCIENCES SOCIETY

*Bull. Malaysian Math. Sc. Soc.* (Second Series) **27** (2004) 73–75

# On Two Methods of Unbiased Estimation in Two-Phase Sampling Using Two Auxiliary Variables

[1]L.N. SAHOO, [2]S.C. SENAPATI AND [3]M. DALABEHERA

[1]Department of Statistics, Utkal University, Bhubaneswar – 751004, India
[2]Department of Statistics, Ravenshaw College, Cuttack – 753003, India
[3]Department of Statistics, Orissa University of Agriculture and Technology, Bhubaneswar – 751003, India
[1]e-mail: nacharya@vsnl..net

**Abstract.** This paper considers two different methods of estimation to generate unbiased estimators of the population mean in the presence of two auxiliary variables for a two-phase sampling procedure.

## 1. Introduction

Let $(y_i, x_i, z_i), i = 1, 2, \cdots, N,$ denote observation vector on the *i*th unit of a trivariate population $\Omega$ in $(y, x, z)$ with mean vector $(\overline{Y}, \overline{X}, \overline{Z})$, such that *y* is the survey variable and $x, z$ are two auxiliary variables. We consider a practical situation where $\overline{X}$ is unknown but $\overline{Z}$ is known exactly, and we seek to estimate $\overline{Y}$ using a two-phase sampling mechanism. Allowing simple random sampling without replacement for sample selection, this scheme is described as follows:

(a) The first phase sample $s'(s' \subset \Omega)$ of fixed size $n'$ is drawn to observe *x* and *z*.
(b) Given $s'$, the second phase sample $s(s \subset s')$ of fixed size *n* is drawn to observe *y* only.

Let $(\bar{y}', \bar{x}', \bar{z}')$ and $(\bar{y}, \bar{x}, \bar{z})$ be the sample mean vectors based on $n'$ and *n* units respectively.

In the above scenarios, basic work on estimation was initiated by Chand (1975) and subsequently studied by several authors producing a huge stock of estimators in survey sampling literature during the last two decades (*cf.* Sahoo and Sahoo [3, 4]). But, the main disadvantage associated with these estimators is that they are seriously biased even if for moderate sized samples. Unfortunately, no attempt has been made so far to construct an unbiased estimator. However, our present paper achieves this objective by creating two generalized unbiased estimators using the concepts introduced in Mickey [2] and Williams [5].

## 2. Mickey-type unbiased estimator

Let us assume that the samples are ordered set of units. For any choice of the first $\alpha$ sample units, remaining $(n'-\alpha)$ units in $s'$ will constitute a first phase sample from the remaining $(N-\alpha)$ units in $\Omega$ and remaining $(n-\alpha)$ units in $s$ will constitute a second phase sample. Let $(\bar{y}_\alpha, \bar{x}_\alpha, \bar{z}_\alpha)$ be the mean vector and $a_1, a_2$ be the functions of the sample values obtained from the first $\alpha$ selections. For given $\alpha$ and $s'$, the expected value of

$$\frac{n\bar{y}-\alpha\bar{y}_\alpha}{n-\alpha} - a_1\left(\frac{n\bar{x}-\alpha\bar{x}_\alpha}{n-\alpha} - \frac{n'\bar{x}'-\alpha\bar{x}_\alpha}{n'-\alpha}\right) \text{ is } \frac{n'\bar{y}'-\alpha\bar{y}_\alpha}{n'-\alpha}$$

and the expected value of

$$\frac{n\bar{y}-\alpha\bar{y}_\alpha}{n-\alpha} - a_1\left(\frac{n\bar{x}-\alpha\bar{x}_\alpha}{n-\alpha} - \frac{n'\bar{x}'-\alpha\bar{x}_\alpha}{n'-\alpha}\right) - a_2\left(\frac{n'\bar{z}'-\alpha\bar{z}_\alpha}{n'-\alpha} - \frac{N\bar{Z}-\alpha\bar{z}_\alpha}{N-\alpha}\right)$$

is $\dfrac{N\bar{Y}-\alpha\bar{y}_\alpha}{N-\alpha}$ .

Hence an unbiased estimator of $\bar{Y}$ is given by

$$t_a = \bar{y} - a_1(\bar{x}-\bar{x}') - a_2(\bar{z}'-\bar{Z}) - \frac{(N-n)\alpha}{N(n-\alpha)}\left[(\bar{y}_\alpha-\bar{y}) - a_1(\bar{x}_\alpha-\bar{x})\right]$$

$$-\frac{(N-n')\alpha}{N(n'-\alpha)}\left[a_1(\bar{x}_\alpha-\bar{x}') - a_2(\bar{z}_\alpha-\bar{z}')\right].$$

## 3. Williams-type unbiased estimator

Assume that $\Omega$ is split at random into $M$ mutually exclusive and exhaustive groups each of size $\frac{n}{m}$ such that $N = \frac{Mn}{n}$. A first phase sample $s'$ of $m'$ groups is taken with equal probabilities without replacement so that $n' = \frac{m'n}{m}$. Then select randomly without replacement $m$ groups from the $m'$ groups of $s'$. This gives second phase sample $s$ of size $n$. For $k$th group of $\frac{n}{m}$ units, $k = 1, 2, \cdots, M$, let $(\bar{y}_k, \bar{x}_k, \bar{z}_k)$ be the mean vector and $b_{1k}, b_{2k}$ be unspecified functions of $y, x$ and $z$. For a given split and a random selection of groups $E(\bar{y}_k, \bar{x}_k, \bar{z}_k) = (\bar{Y}, \bar{X}, \bar{Z})$ and $\bar{y} = \frac{1}{m}\sum_{k=1}^{m}\bar{y}_k$ , $\bar{x}' = \frac{1}{m'}\sum_{k=1}^{m'}\bar{x}_k$ ,

$\bar{Z} = \frac{1}{M} \sum_{k=1}^{M} \bar{z}_k$ etc.

To construct an Williams-type unbiased estimator, we first consider the estimator $\bar{y} + \bar{b}_1 (\bar{x}' - \bar{x}) + \bar{b}_2 (\bar{Z} - \bar{z}')$ and its bias $-[Cov(\bar{b}_1, \bar{x}) - Cov(\bar{b}_1, \bar{x}') + Cov(\bar{b}_2, \bar{z}')]$, where $\bar{b}_1 = \frac{1}{m} \sum_{k=1}^{m} b_{1k}$ and $\bar{b}_2 = \frac{1}{m} \sum_{k=1}^{m} b_{2k}$. Then adjusting the estimator by a sample estimate of the needed bias correction, an unbiased estimator of $\bar{Y}$ may be defined by

$$t_b = \bar{y} + \bar{b}_1 (\bar{x}' - \bar{x}) + \bar{b}_2 (\bar{Z} - \bar{z}') + \left(1 - \frac{n}{n'}\right) \frac{1}{m(m-1)} \sum_{k=1}^{m} b_{1k} (\bar{x}_k - \bar{x})$$
$$+ \left(1 - \frac{n'}{N}\right) \frac{1}{m'(m-1)} \sum_{k=1}^{m} b_{2k} (\bar{z}_k - \bar{z}) .$$

## 4. Conclusions

Unbiasedness of the estimators $t_a$ and $t_b$ can be displayed for any defined form of their coefficients *i.e.*, $a_1, a_2$ for $t_a$ and $b_{1k}, b_{2k}$ for $t_b$. Thus, these estimators define systems of unbiased estimators for $\bar{Y}$ .

Discussion of the paper is somewhat theoretical in nature but encouraging for the future development in the construction of unbiased estimators for the situation under consideration. The methodologies presented here will be useful in a circumstance where the statistician can reasonably choose the value of $\alpha$ for $t_a$, values of $m$ and $M$ for $t_b$, and the coefficients for the estimators. Complexity of an estimator can also be avoided with the judicious choice of the coefficients.

## References


1. L. Chand, *Some Ratio-type Estimators Based on Two or More Auxiliary Variables*, Unpublished Ph.D. dissertation, Iowa State University, Ames, Iowa, 1975.
2. M.R. Mickey, Some finite population unbiased ratio and regression estimators, *Jour. Amer. Stat. Assoc.* **54** (1959), 594–612.
3. J. Sahoo and L.N. Sahoo, A class of estimators in two-phase sampling using two auxiliary variables, *Jour. Ind. Stat. Assoc* **31** (1993), 107–114.
4. J. Sahoo and L.N. Sahoo, An alternative class of estimators in double sampling procedures, *Calcutta Stat. Assoc. Bull.* **49** (1999), 79–83.
5. W.H. Williams, Generating unbiased ratio and regression estimators, *Biometrics* **17** (1961), 267–274.


Keywords and phrases: Auxiliary variable, two-phase sampling, unbiased estimator.