ISSN 1684 – 8403
Journal of Statistics
Vol: 13, No.1 (2006)

A New Symmetrized Estimator of Population Total in Unequal Probability Sampling

Ayesha Siddique[1]
Muhammad Qaiser Shahbaz

Abstract

A new symmetrized estimator of population total has been obtained for use in unequal probability sampling without replacement. Empirical study of the new estimator is done. It is found that the new estimator perform reasonably well as compared to the Murthy (1957) estimator.

Key words

Unequal probability sampling, Murthy Estimator.

1. Introduction

Horvitz and Thompson (1952) originate the theory of unequal probability sampling without replacement when they proposed following estimator of population total:

$$y'_{HT} = \sum_{i=1}^{n} \frac{y_i}{\pi_i} \quad , \qquad (1.1)$$

where π_i is probability of inclusion of i-th unit in the sample. Horvitz and Thompson gave following variance formula for estimator (1.1).

$$V\left(y'_{HT}\right) = \sum_{i=1}^{N} \frac{\left(1-\pi_i\right)}{\pi_i} \; Y_i^2 + \sum_{i=1}^{N}\sum_{\substack{j=1 \\ j\neq i}} \frac{\left(\pi_{ij}-\pi_i\pi_j\right)}{\pi_i\pi_j} Y_iY_j \, , \qquad (1.2)$$

an alternative expression, for fixed n, given by Sen (1953) and independently by Yates and Grundy (1953), is:

[1] Department of Statistics, GC University Lahore
Received Jan 2007, Revised April 2007

$$V\left(y_{HT}^{/}\right)=\sum_{i=1}^{N}\sum_{\substack{j=1\\j>i}}\left(\pi_i\pi_j-\pi_{ij}\right)\left(\frac{Y_i}{\pi_i}-\frac{Y_j}{\pi_j}\right)^2. \quad (1.3)$$

Number of selection procedures are derived for use with the Horvitz–Thompson estimator since its emergence.

Raj (1956) developed following ordered estimator for population total:

$$t_{mean}=\frac{1}{n}\sum_{r=1}^{n}t_r \quad , \qquad (1.4)$$

where $t_1=\frac{y_1}{p_1}$ and $t_r=\sum_{i=1}^{r-1}y_i+\frac{y_r}{p_r}\left(1-\sum_{i=1}^{r-1}p_i\right)$ *for* $r>1$.

(1.5)

The Raj (1956) estimator for a sample of size 2 is given as:

$$t_{mean}=\frac{1}{2}\left[\frac{y_i}{p_i}\left(1+p_i\right)+\frac{y_i}{p_i}\left(1-p_j\right)\right]. \quad (1.6)$$

the sampling variance of estimator given in (1.6) is given as:

$$Var\left(t_{mean}\right)=\frac{1}{8}\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}P_iP_j\left(2-P_i-P_j\right)\left(\frac{Y_i}{P_i}-\frac{Y_j}{P_j}\right)^2. \quad (1.7)$$

Murthy (1957) uses the idea of sufficiency to overcome the defect of Raj estimator. He symmetrized the Raj's estimator to produce an un-ordered estimator. The estimator proposed by Murthy has general form:

$$t_{symm}=\frac{1}{P(s)}\sum_{i=1}^{n}P\left(s\mid i\right)y_i \quad (1.8)$$

where $P\left(s\mid i\right)$ is the probability of obtaining a sample "s" given that ith unit has been already selected and $P(s)$ is the probability of obtaining a sample "s"

The Murthy (1957) estimator for a sample of size 2 is given as:

$$t_{symm}=\frac{1}{2-p_i-p_j}\left[\frac{y_i}{p_i}\left(1-p_j\right)+\frac{y_j}{p_j}\left(1-p_i\right)\right]. \quad (1.9)$$

The sampling variance for estimator given in (1.9) is given as:

$$Var\left(t_{symm}\right)=\frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}\frac{P_iP_j\left(1-P_i-P_j\right)}{2-P_i-P_j}\cdot\left(\frac{Y_i}{P_i}-\frac{Y_j}{P_j}\right)^2 . \quad (1.10)$$

2. The New Symmetrized Estimator

Murthy (1957) symmetrized the Raj (1956) estimator to obtain estimator (1.9) for estimation of population total. To obtain the new symmetrized estimator, we have considered following modified versions of the Raj's estimator given by Samiuddin and Kattan (1991):

$$\left.\begin{aligned}&T_1=y_1+\frac{y_2}{p_2}\left(1-p_2\right),\\&T_2=y_2+\frac{y_1}{p_1}\left(1-p_1\right)\left(k-\frac{p_2}{1-p_2}\right)\qquad with\quad k=\sum_{i=1}^{N}\frac{p_i}{1-p_i}\end{aligned}\right\}\quad (2.1)$$

Now using the idea of Murthy, the new symmetrized estimator is obtained below.

We first find the average of T_1 and T_2 to obtain the following estimator:

$$\begin{aligned}T_m&=\frac{1}{2}\left(T_1+T_2\right)\\&=\frac{1}{2}\left[y_i+\frac{y_j}{p_j}\left(1-p_i\right)+y_j+\frac{y_i}{p_i}\left(1-p_i\right)\left(k-\frac{p_j}{1-p_j}\right)\right]\\&=\frac{1}{2}\left[\frac{y_i}{p_i}\left\{p_i+\left(1-p_i\right)\left(k-\frac{p_j}{1-p_j}\right)\right\}+\frac{y_j}{p_j}\left(1-p_i+p_j\right)\right]\end{aligned}\quad (2.2)$$

To symmetrize this estimator we formulate following two estimators based on order of selection of units:

$$\left.\begin{aligned}&T_m\left(i,j\right)=\frac{1}{2}\left[\frac{y_i}{p_i}\left\{p_i+\left(1-p_i\right)\left(k-\frac{p_j}{1-p_j}\right)\right\}+\frac{y_j}{p_j}\left(1-p_i+p_j\right)\right]\\&T_m\left(j,i\right)=\frac{1}{2}\left[\frac{y_j}{p_j}\left\{p_j+\left(1-p_j\right)\left(k-\frac{p_i}{1-p_i}\right)\right\}+\frac{y_i}{p_i}\left(1-p_j+p_i\right)\right]\end{aligned}\right\}\quad (2.3)$$

Now the symmetrized estimator is obtained below:

$$T_{symm} = \frac{T_m(i,j)P(i,j) + T_m(j,i)P(j,i)}{P(i,j)+P(j,i)} \tag{2.4}$$

where

$$P(i,j) = p_i p_{j|i} = \frac{p_i p_j}{1-p_i}, \qquad P(j,i) = p_j p_{i|j} = \frac{p_j p_i}{1-p_j} \tag{2.5}$$

substituting values from (2.3) and (2.5) in (2.4) we have:

$$T_{symm} = \left[\frac{p_i p_j}{1-p_i} + \frac{p_i p_j}{1-p_j}\right]^{-1} \left[\frac{p_i p_j}{1-p_i}\left\{\frac{1}{2}\left[\frac{y_i}{p_i}\left\{p_i + (1-p_i)\left(k - \frac{p_j}{1-p_j}\right)\right\} + \frac{y_j}{p_j}\left(1-p_i+p_j\right)\right]\right\}\right.$$

$$\left. + \frac{p_i p_j}{1-p_j}\left\{\frac{1}{2}\left[\frac{y_j}{p_j}\left\{p_j + (1-p_j)\left(k - \frac{p_i}{1-p_i}\right)\right\} + \frac{y_i}{p_i}\left(1-p_j+p_i\right)\right]\right\}\right]$$

After slight algebraic manipulation we get following new symmetrized estimator:

$$\left.\begin{aligned} T_{symm} = &\frac{(y_i/p_i)\left[(1-p_j)\{1+k(1-p_i)\} + (1-p_i)(p_i-p_j)\right]}{2(2-p_i-p_j)} \\ &+ \frac{(y_j/p_j)\left[(1-p_i)\{1+k(1-p_j)\} + (1-p_j)(p_j-p_i)\right]}{2(2-p_i-p_j)} \end{aligned}\right\} \tag{2.6}$$

The design based variance of this symmetrized estimator is derived below:

Now $Var(T_{symm}) = E(T_{symm} - Y)^2$,

$$Var(T_{symm}) = \sum_{i=1}^{N}\sum_{\substack{j=1 \\ j\neq i}} T_{symm}^2 P(S) - Y^2. \tag{2.7}$$

Substituting the value of T_{symm} from (2.6) in (2.7), and after some simplification, the design based variance of new symmetrized estimator is given as:

$$Var(T_{symm}) = \frac{1}{2}\sum_{i=1}^{N}\sum_{\substack{j=1 \\ j\neq i}} \frac{p_i p_j}{4(1-p_i)(1-p_j)(2-p_i-p_j)}\left[\frac{y_i^2}{p_i^2}A_{ij} + \frac{y_j^2}{p_j^2}B_{ij} - 2\frac{y_i y_j}{p_i p_j}C_{ij}\right] \tag{2.8}$$

where

$$A_{ij} = \left[\left\{(1-p_j)\left[1+k(1-p_i)\right]+(1-p_i)(p_i-p_j)\right\}^2 - 4p_i(1-p_j)(2-p_i-p_j)\right]$$

$$B_{ij} = \left[\left\{(1-p_i)\left[1+k(1-p_j)\right]+(1-p_j)(p_j-p_i)\right\}^2 - 4p_j(1-p_i)(2-p_i-p_j)\right]$$

$$C_{ij} = \Big[4(1-p_i)(1-p_j)(2-p_i-p_j) - \left\{(1-p_j)\left[1+k(1-p_i)\right]+(1-p_i)(p_i-p_j)\right\}$$
$$\left\{(1-p_i)\left[1+k(1-p_j)\right]+(1-p_j)(p_j-p_i)\right\}\Big]\Big]$$

The empirical study of (2.6) is given in section 3 below.

3. Empirical study

In this section, we have carried out the empirical study of the new symmetrized estimator. For this empirical study, we have selected 15 natural populations. For comparative purpose sampling variance of Murthy (1957) estimator and the new symmetrized estimator has been calculated from these fifteen populations. The results of these sampling variances are given in following table:

Pop. No.	**Murthy**	**New Estimator**
1.	275.093	247.000
2.	464573.800	463880.200
3.	132624.100	129719.800
4.	8528.188	8256.219
5.	29.549	28.756
6.	1228.399	1210.221
7.	369817.700	363676.200
8.	42606180	42108768
9.	23619.030	23430.370
10.	597.566	434.339

From above table, we can readily see that the New Symmetrized Estimator has far better performance than the Murthy (1957) estimator.

References:

1. Horvitz, D. G. and Thompson, D. J. (1952) "A generalization of sampling without replacement from a finite universe", *J. Amer. Stat. Assoc*. 47, 663 – 685.
2. Murthy, M. N. (1957) "Ordered and unordered estimators in sampling without replacement", *Sankhya*, 18, 379 – 390.
3. Raj, D. (1956) "Some estimators in sampling with varying probabilities without replacement", *J. Amer. Stat. Assoc*. 51, 269 – 284.
4. Samiuddin, M., Kattan, A. K. A., Hanif, M. and Asad, H. (1992) "Some remarks on models, sampling schemes and estimators in unequal probability sampling", *Pak. J. Stat.*, 8(1), A, 1 – 18
5. Sen, A. R. (1953). "On the estimate of the variance in sampling with varying probabilities" *J. Indian Soc. Agri. Statist. 5, 119 – 127.*
6. Yates, F. and Grundy, P. M. (1953) "Selection without replacement from within strata with probability proportional to size", *J. Roy. Stat. Soc*., B, 15, 153 – 161.