International Journal of Applied Mathematics
& Statistical Sciences (IJAMSS)
ISSN(P): 2319-3972; ISSN(E): 2319-3980
Vol. 2, Issue 5, Nov 2013, 125-142

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

AN EXACT EXPRESSION FOR THE MEAN SQUARE ERROR OF RATIO ESTIMATOR, REGRESSION ESTIMATOR AND GENERALIZED REGRESSION ESTIMATOR IN FINITE POPULATION SAMPLING

ARUN KUMAR ADHIKARY

Bayesian and Interdisciplinary Research Unit, Indian Statistical Institute, Kolkata, West Bengal, India

ABSTRACT

Following Rao (1979) an attempt has been made to derive an exact expression for the mean square error of ratio estimator, regression estimator, generalized regression estimator, separate ratio and regression estimator and combined ratio and regression estimator in stratified random sampling and also an exact expression for an unbiased estimator of their mean square error. Noting that Rao's (1979) procedure fails to derive an exact expression for the mean square error of product estimator of the population total, an alternative procedure is suggested to get an exact expression for the variance of a homogeneous linear unbiased estimator of the population total and also an exact expression for an unbiased estimator of the variance of the estimator. This procedure is illustrated in case of Horvitz-Thompson (1952) estimator, Hansen-Hurwitz (1943) estimator based on PPSWR sampling, Murthy's (1957) unordered estimator based on PPSWOR sampling, ratio estimator based on Lahiri (1951), Midzuno (1952) and Sen's (1953) sampling scheme and Hartley – Ross (1954) unbiased ratio type estimator based on SRSWOR sampling scheme.

KEYWORDS: Generalized Regression Estimator, Homogeneous Linear Unbiased Estimator, Mean Square Error, Ratio Estimator, Regression Estimator, Separate and Combined Ratio Estimator, Separate and Combined Regression Estimator

1. INTRODUCTION

Let $U = (1,2,\dots,N)$denote a finite population of size N and let y_i denote the value of a study variable y assumed on the ith unit of the population. Our problem is to estimate the population total $Y = \sum_{i=1}^{N} y_i$based on a sample s drawn from a population of size N with a probability p(s).

A homogeneous linear estimator of the population total Y is given by

$$\hat{Y} = \sum_{i=1}^{N} d_{si}\, y_i$$

where d_{si}'s are independent of y_i's but may depend on x_i's where x_i is the value of an auxiliary variable x highly correlated with the study variable y assumed on the ith unit and $d_{si} = 0$ if i does not belong to s or s does not contain i.

According to Rao (1979), if MSE$(\hat{Y}) = 0$ for for some choice of $y_i$as $y_i = cw_i$ where $c \neq 0$ and w_i's are some known constants, the MSE$(\hat{Y})$ can be written as

$$\text{MSE}(\hat{Y}) = -\sum_{i<j=1}^{N}\sum_{j=1}^{N} d_{ij}\, w_i w_j \left(\frac{y_i}{w_i} - \frac{y_j}{w_j}\right)^2$$

where $d_{ij} = E_p\left[(d_{si} - 1)(d_{sj} - 1)\right]$

$$= E_p(d_{si} d_{sj}) - E_p(d_{si}) - E_p(d_{sj}) + 1$$

$$= \sum_{s \ni i,j} d_{si}\, d_{sj}\, p(s) - \sum_{s \ni i} d_{si}\, p(s) - \sum_{s \ni j} d_{sj}\, p(s) + 1.$$

An unbiased estimator of MSE$(\hat{Y})$ is given by

$$\widehat{MSE}(\hat{Y}) = -\sum_{i<j=1}^{N}\sum_{j=1}^{N} d_{ij}\ (s) w_i w_j \left(\frac{y_i}{w_i} - \frac{y_j}{w_j}\right)^2$$

where $d_{ij}\,(s) = 0$ if s does not contain i and j and $d_{ij}\,(s)$ is such that

$E_p[d_{ij}\,(s)] = d_{ij}$ or $\sum_{s\ni i,j} d_{ij}\ (s)p(s) = d_{ij}$.

The several choices of $d_{ij}\,(s)$ have been suggested by Rao(1979) but they are all applicable in case of a homogeneous linear unbiased estimator $\hat{Y}$ in which case $\sum_{s\ni i} d_{si}\, p(s) = 1 \forall i$ and as a consequence d_{ij} reduces to $d_{ij} = \sum_{s\ni i,j} d_{si}\, d_{sj}\, p(s) - 1$.

In case $\hat{Y}$ corresponds to a biased estimator, the only possible choice of $d_{ij}\,(s)$ is $d_{ij}\,(s) = \frac{d_{ij}}{\pi_{ij}}$ where $\pi_{ij} = \sum_{\square\ni\,|\,,j} p(s)$is the second order inclusion probability so that $\sum_{s\ni i,j} d_{ij}\ (s)p(s) = d_{ij}$.

2. MEAN SQUARE ERROR OF RATIO ESTIMATOR

The ratio estimator of the population total Y is given by

$$\widehat{Y_R} = \frac{\bar{y}}{\bar{x}} X$$

where $\bar{y} = \frac{1}{n}\sum_{i=1}^{\text{D}} y_i$ and $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ are the sample means of y and x respectively based on a sample of size n drawn from a population of size N by SRSWOR and $X = \sum_{i=1}^{N} x_i$ is the population total of the auxiliary variable x.

Now $\widehat{Y_R}$ can be written as

$$\widehat{Y_R} = \sum_{i=1}^{N} d_{si} y_i$$

where $d_{si} = \frac{X}{\sum_{i\in s} x_i} = \frac{1}{\sum_{i\in s} p_i} = \frac{1}{p_s}$ if $i \in s$ or $s \ni i$ where $p_i = \frac{x_i}{X}$ and $p_s = \sum_{i\in s} p_i$ and $d_{si} = 0$ otherwise.

We may note that MSE$(\widehat{Y_{\langle}}) = 0$ if $y_i = cx_i\ \forall i$ and as a consequence MSE$(\widehat{Y_R})$ can be written as

$$\text{MSE}(\widehat{Y_R}) = -\sum_{i<j=1}^{N}\sum_{j=1}^{N} d_{ij}\ x_i x_j \left(\frac{y_i}{x_i} - \frac{y_j}{x_j}\right)^2$$

where $d_{ij} = \sum_{s\ni i,j} d_{si}\, d_{sj}\, p(s) - \sum_{s\ni i} d_{si}\ p(s) - \sum_{s\ni j} d_{sj}\ p(s) + 1$

$$= \sum_{s\ni i,j} \frac{1}{p_s^2}\frac{1}{\binom{N}{n}} - \sum_{s\ni i}\frac{1}{p_s}\frac{1}{\binom{N}{n}} - \sum_{s\ni j}\frac{1}{p_s}\frac{1}{\binom{N}{n}} + 1$$

An unbiased estimator of MSE$(\widehat{Y_R})$ is given by

$$\widehat{MSE}(\widehat{Y_R}) = -\sum_{i<j=1}^{N}\sum_{j=1}^{N} d_{j}\ (s) x_i x_j \left(\frac{y_i}{x_i} - \frac{y_j}{x_j}\right)^2$$

where $d_{ij}\,(s) = 0$ if s does not contain i and j and $d_{ij}\,(s) = \frac{d_{ij}}{\pi_{ij}}$ if $i, j \in s$ where $\pi_{ij} = \frac{n(n-1)}{N(N-1)}$.

3. MEAN SQUARE ERROR OF REGRESSION ESTIMATOR

The regression estimator of the population total is given by

$$\widehat{Y_{lr}} = N[\bar{y} + b(\bar{X} - \bar{x})]$$

Where $\bar{y}, \bar{x}$ are the sample means of y and x respectively based on a sample of size n drawn from a population of size N by SRSWOR and $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} x_i$ is the population mean of the auxiliary variable x.

Here b is the sample regression coefficient of y on x and is given by

$$b = \frac{\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})}{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{\sum_{i-1}^{n} y_i(x_i-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

Now $\widehat{Y_{lr}}$ can be written as $\widehat{Y_{lr}} = \sum_{i=1}^{N} d_{si}\, y_i$

Where $d_{si} = 0$ if s does not contain i or i does not belong to s and

$d_{si} = N\left[\frac{1}{n} + \frac{(x_i-\bar{x})}{\sum_{i\in s}(x_i-\bar{x})^2}(\bar{X}-\bar{x})\right]$ if $s \ni i$ or $i \in s$.

We may note that $\text{MSE}(\widehat{Y_{lr}}) = 0$ if $y_i = cx_i\ \forall i$ and as a consequence $\text{MSE}(\widehat{Y_{lr}})$ can be written as

$$\text{MSE}(\widehat{Y_{lr}}) = -\sum_{i<j=1}^{N}\sum_{j=1}^{N} d_{ij}\ \ x_i x_j\left(\frac{y_i}{x_i} - \frac{y_j}{x_j}\right)^2$$

Where $d_{ij} = \sum_{s\ni i,j} d_{si}\, d_{sj}\, p(s) - \sum_{s\ni i} d_{si}\, p(s) - \sum_{s\ni j} d_{sj}\ p(s)$ +1 where $p(s) = \frac{1}{\binom{N}{n}}$.

An unbiased estimator of $\text{MSE}(\widehat{Y_{lr}})$ is given by

$$\widehat{MSE}\ \langle\widehat{Y_{lr}}\rangle = -\sum_{i<j=1}^{N}\sum_{j=1}^{N} d_{ij}\ (s) x_i x_j\left(\frac{y_i}{x_i} - \frac{y_j}{x_j}\right)^2$$

Where $d_{ij}\,(s) = 0$ if s does not contain i and j and $d_{ij}\,(s) = \frac{d_{ij}}{\pi_{ij}}$ if $i, j \in s$ where $\pi_{ij} = \frac{n(n-1)}{N(N-1)}$.

4. MEAN SQUARE ERROR OF GENERALIZED REGRESSION ESTIMATOR

Let a population of size N be divided into D non-overlapping domains U_d such that $U_d \cap U_{d'} = \emptyset$ for $d \neq d'$ and $\bigcup_{d=1}^{D} U_d = U$.

Let $Y_d = \sum_{i\in U_d} y_i$ and $X_d = \sum_{i\in U_d} x_i$ denote the population domain totals of y and x respectively. Let a sample s be drawn from the population with a probability (s).

Let $s_d = s \cap U_d$ denote the part of the sample s coming from U_d. Then the generalized regression estimator of Y_d is given by

$$\widehat{Y_d} = X_d\widehat{\beta_Q} + \sum_{i\in s_d}\frac{e_i}{\pi_i}$$

Where $\widehat{\beta_Q} = \frac{\sum_{i\in s} y_i x_i Q_i}{\sum_{i\in s} x_i^2 Q_i}$, Q_i's being arbitrarily assignable positive constants and $e_i = y_i - \widehat{\beta_Q} x_i$, $\pi_i = \sum_{s\ni i} p(s)$.

Then following Cassel, S$\ddot{a}rndal$ and Wretman (1976), Särndal (1980, 1982), Chaudhuri and Adhikary (1995, 1998) and Adhikary (2000, 2005), $\widehat{Y_d}$ can be written as

$$\widehat{Y_d} = \sum_{i\in s}\frac{y_i}{\pi_i} g_{sdi}$$

Where $g_{sdi} = I_{di} + \left(X_d - \sum_{i\in s_d}\frac{x_i}{\pi_i}\right)\frac{\pi_i x_i Q_i}{\sum_{i\in s} x_i^2 Q_i}$

Where $I_{di} = 1$ if $i \in U_d$

$= 0$ otherwise.

We may note that MSE$(\widehat{Y_d}) = 0$ if $y_i = cx_i\ \forall i$ and as a consequence MSE$(\widehat{Y_d})$can be written as

$$\text{MSE}(\widehat{Y_d}) = -\sum_{i<j=1}^{N}\sum_{j=1}^{N} d_{ij}\ x_i x_j \left(\frac{y_i}{x_i} - \frac{y_j}{x_j}\right)^2$$

Where $d_{ij} = \sum_{s\ni i,j} d_{si}\, d_{sj}\, p(s) - \sum_{s\ni i} d_{si}\, p(s) - \sum_{s\ni j} d_{sj}\ p(s) + 1$

Where $d_{si} = 0$ if s does not contain i or i does not belong to s and

$d_{si} = \frac{g_{sdi}}{\pi_i}$ if $\in s$.

An unbiased estimator of of MSE$(\widehat{Y_d})$ is given by

$$\widehat{MSE}(\widehat{Y_d}) = -\sum_{i<j=1}^{N}\sum_{j=1}^{N} d_{ij}\ (s) x_i x_j \left(\frac{y_i}{x_i} - \frac{y_j}{x_j}\right)^2$$

Where $d_{ij}\,(s) = 0$ if s does not contain i and j and $d_{ij}\,(s) = \frac{d_{ij}}{\pi_{ij}}$ if $i, j \in s$.

5. MEAN SQUARE ERROR OF SEPARATE RATIO ESTIMATOR

Let a population of size N be divided into L strata, the hth stratum consisting of $N_h$units so that $\sum_{h=1}^{L} N_h = N$ and let y_{hi} and x_{hi}denote respectively the value of the study variable y and the auxiliary variable x assumed on the ith unit of the hth stratum, i=1,2,…,N_h, $h = 1,2, \dots, L$.

Let a sample $s_h$of size n_h be drawn from the hth stratum by SRSWOR. Then the separate ratio estimator of the population total is given by

$$\widehat{Y_{RS}} = \sum_{h=1}^{L} \frac{\overline{y_h}}{\overline{x_h}} X_h$$

Where $\overline{y_h} = \frac{1}{n_h}\sum_{i=1}^{n_h} y_{hi}$ and $\overline{x_h} = \frac{1}{n_h}\sum_{i=1}^{n_h} x_{hi}$ are the sample means of y and x respectively for the hth stratum and $X_h = \sum_{i=1}^{N_h} x_{hi}$ is the population total of x for the hth stratum, h=1,2,…,L.

Now $\widehat{Y_{RS}}$ can be written as

$$\widehat{Y_{RS}} = \sum_{h=1}^{L}\sum_{i=1}^{N_h} d_{hsi}\ y_{hi}$$

WHERE $d_{hsi} = 0$ if i does not belong to $s_h$or s_h does not contain i and

$d_{h振i} = \frac{X_h}{\sum_{i\in s_h} x_{hi}} = \frac{1}{\sum_{i\in s_h} p_{hi}} = \frac{1}{p_{s_h}}$, say, if $i \in s_h$ where $p_{hi} = \frac{x_{hi}}{X_h}$ and $p_{s_h} = \sum_{i\in s_h} p_{hi}$.

We may note that MSE$(\widehat{Y_{RS}}) = 0$ if $y_{hi} = cx_{hi}\ \forall i$ and as a consequence MSE$(\widehat{Y_{RS}})$ can be written as

$$\text{MSE}(\widehat{Y_{RS}}) = -\sum_{h=1}^{L}\sum_{i<j=1}^{N_h}\sum_{j=1}^{N_h} d_{hij}\ x_{hi} x_{hj} \left(\frac{y_{hi}}{x_{hi}} - \frac{y_{hj}}{x_{hj}}\right)^2$$

Where $d_{hij} = \sum_{s_h\ni i,j} d_{hsi}\, d_{hsj}\, p(s) - \sum_{s_h\ni i} d_{hsi}\, p(s) - \sum_{s_h\ni j} d_{hsj}\ p(s)$ +1.

An unbiased estimator of MSE$(\widehat{Y_{RS}})$ is given by

$$\widehat{MSE}(\widehat{Y_{RS}}) = -\sum_{h=1}^{L}\sum_{i<j=1}^{N_h}\sum_{j=1}^{N_h} d_{hij}\ (s_h) x_{hi} x_{hj} \left(\frac{y_{hi}}{x_{hi}} - \frac{y_{hj}}{x_{hj}}\right)^2$$

Where $d_{hij}\,(s_h) = 0$ if i and j do not belong to s_h and

$d_{hij}\,(s_h) = \frac{d_{hij}}{\pi_{hij}}$ if $i, j \in s_h$ where π_{hij} is the inclusion probability of a pair of units i and j from the hth stratum and is given by $\pi_{hij} = \frac{n_h(n_h-1)}{N_h(N_h-1)}$.

6. MEAN SQUARE ERROR OF COMBINED RATIO ESTIMATOR

The combined ratio estimator of the population total Y is given by

$$\widehat{Y_{Rc}} = \frac{\overline{y_{st}}}{\overline{x_{st}}} X$$

Where $\overline{y_{st}} = \sum_{h=1}^{L} W_h\, \overline{y_h}, \overline{x_{st}} = \sum_{h=1}^{L} W_h\, \overline{x_h}$ and $X = \sum_{h=1}^{L} X_h$ where $W_h = \frac{N_h}{N}$, $h = 1,2, \dots, L$.

Now $\widehat{Y_{RC}}$ can be written as

$$\widehat{Y_{RC}} = \sum_{h=1}^{L} \sum_{i=1}^{N_h} d_{hsi}\, y_{hi}$$

Where $d_{hsi} = 0$ if i does not belong to s_h or s_h does not contain i and

$$d_{hsi} = \frac{N_h}{n_h} . \frac{X}{\sum_{h=1}^{L} N_h \overline{x_h}} \text{ if } i \in s_h.$$

We may note that $MSE(\widehat{Y_{RC}}) = 0$ if $y_{hi} = c x_{hi}\ \forall i$ and as a consequence $MSE(\widehat{Y_{RC}})$ can be written as

$$MSE(\widehat{Y_{RC}}) = -\sum_{h=1}^{L} \sum_{i<j=1}^{N_h} \sum_{j=1}^{N_h} d_{hij}\, x_{hi} x_{hj} \left(\frac{y_{hi}}{x_{hi}} - \frac{y_{hj}}{x_{hj}}\right)^2$$

Where $d_{hij} = \sum_{s_h \ni i,j} d_{hsi}\, d_{hsj}\, p(s_h) - \sum_{s_h \ni i} d_{hsi}\, p(s_h) - \sum_{s_h \ni j} d_{hs,}\, p(s_h) + 1$

Where $p\,(s_h) = \frac{1}{\binom{N_h}{n_h}}$.

An unbiased estimator of $MSE(\widehat{Y_{RC}})$ is given by

$$\widehat{MSE}(\widehat{Y_{RC}}) = -\sum_{h=1}^{L} \sum_{i<j=1}^{N_h} \sum_{j=1}^{N_h} d_{hij}\ (s_h) x_{hi} x_{hj} \left(\frac{y_{hi}}{x_{hi}} - \frac{y_{hj}}{x_{hj}}\right)^2$$

Where $d_{hij}\,(s_h) = 0$ if s_h does not contain i and j and $d_{hij}\,(s_h) = \frac{d_{hij}}{\pi_{hij}}$ if $i, j \in s_h$, where π_{hij} is the inclusion probability of a pair of units i and j from the hth stratum and is given by $\pi_{hij} = \frac{n_h(n_h-1)}{N_h(N_h-1)}, h = 1,2, \dots, L$.

7. MEAN SQUARE ERROR OF SEPARATE REGRESSION ESTIMATOR

The separate regression estimator of the population total is given by

$$\widehat{Y_{lrs}} = \sum_{h=1}^{L} N_h\, [\overline{y_h} + b_h(\overline{X_h} - \overline{x_h})]$$

Where $b_h = \frac{\sum_{i=1}^{n_h}(y_{hi}-\overline{y_h})(x_{hi}-\overline{x_h}))}{\sum_{i=1}^{n_h}(x_{hi}-\overline{x_h})^2} = \frac{\sum_{i=1}^{n_h} y_{hi}(x_{hi}-\overline{x_h})}{\sum_{i=1}^{n_h}(x_{hi}-\overline{x_h})^2}$ is the sample regression coefficient of y on x in the hth stratum and $\overline{X_h} = \frac{X_h}{N_h}$ is the population mean of the auxiliary variable x in the hth stratum, h=1,2,…,L.

Now $\widehat{Y_{lrs}}$ can be written as

$$\widehat{Y_{lrs}} = \sum_{h=1}^{L} \sum_{i=1}^{N_h} d_{hsi}\, y_{hi}$$

Where $d_{hsi} = 0$ if i does not belong to s_h or s_h does not contain i and

$$d_{hsi} = \frac{N_h}{n_h}\left[1 + \frac{(x_{hi}-\overline{x_h})}{\sum_{i\in s_h}(x_{hi}-\overline{x_h})^2}(\overline{X_h} - \overline{x_h})\right] \text{ if} i \in s_h.$$

We may note that $MSE(\widehat{Y_{lrs}}) = 0$ if $y_{hi} = cx_{hi}$ $\forall i$ and as a consequence $MSE(\widehat{Y_{lrs}})$ can be written as

$$MSE(\widehat{Y_{lrs}}) = -\sum_{h=1}^{L}\sum_{i<j=1}^{N_h}\sum_{j=1}^{N_h} d_{hij}\, x_{hi}x_{hj}\left(\frac{y_{hi}}{x_{hi}} - \frac{y_{hj}}{x_{hj}}\right)^2$$

Where $d_{hij} = \sum_{s_h \ni i,j} d_{hsi}\, d_{hsj}\, p(s_h) - \sum_{s_h\ni i} d_{hsi}\, p(s_h) - \sum_{s_h \ni j} d_{hsj}\, p(s_h) + 1$

An unbiased estimator of$MSE(\widehat{Y_{lrs}})$ is given by

$$\widehat{M\text{栩く}}\,(\widehat{Y_{lrs}}) = -\sum_{h=1}^{L}\sum_{i<j=1}^{N_h}\sum_{j=1}^{N_h} d_{hij}\,(s_h)x_{hi}x_{hj}\left(\frac{y_{hi}}{x_{hi}} - \frac{y_{hj}}{x_{hj}}\right)^2$$

Where $d_{hij}\,(s_h) = 0$ if s_h does not contain i, j or i, j do not belong to s_h and

$d_{hij}\,(s_h) = \frac{d_{hij}}{ð_{hij}}$, where $ð_{hij}$ is the inclusion probability of a pair of units i and j from the hth stratum and is given by $ð_{hij} = \frac{n_h(n_h-1)}{N_h(N_h-1)}$, $h = 1,2, \ldots, L$.

8. MEAN SQUARE ERROR OF COMBINED REGRESSION ESTIMATOR

The combined regression estimator of the population total is given by

$$\widehat{Y_{lrc}} = N[\overline{y_{st}} + b(\bar{X} - \overline{x_{st}})]$$

Where $b = \frac{\sum_{h=1}^{L}\sum_{i=1}^{n_h}(y_{hi}-\overline{y_h})(x_{hi}-\overline{x_h})}{\sum_{h=1}^{L}\sum_{i=1}^{n_h}(x_{hi}-\overline{x_h})^2} = \frac{\sum_{h=1}^{L}\sum_{i=1}^{\text{獨}} y_{hi}(x_{hi}-\overline{x_h})}{\sum_{h=1}^{L}\sum_{i=1}^{n_h}(x_{hi}-\overline{x_h})^2}$ is the sample regression coefficient of y on x and $\bar{X} = \sum_{h=1}^{L} W_h\, \overline{X_h}, W_h = \frac{N_h}{N}, h = 1,2, \ldots, L$

Now $\widehat{Y_{lrc}}$ can be written as

$$\widehat{Y_{lrc}} = \sum_{h=1}^{L}\sum_{i=1}^{N_h} d_{hsi}\, y_{hi}$$

Where $d_{hsi} = 0$ if i does not belong to s_h or s_h does not contain i and

$$d_{hsi} = N\left[\frac{W_h}{n_h} + \frac{x_{hi}-\overline{x_h}}{\sum_{h=1}^{L}\sum_{i=1}^{n_h}(x_{hi}-\overline{x_h})^2}\right] \text{ if } i \in s_h.$$

We may note that $MSE(\widehat{Y_{lrc}}) = 0$ if $y_{hi} = cx_{hi}$ $\forall i$ and as a consequence $MSE(\widehat{Y_{lrc}})$ can be written as

$$MSE(\widehat{Y_{lrc}}) = -\sum_{h=1}^{L}\sum_{i<j=1}^{N_h}\sum_{j=1}^{N_h} d_{hij}\, x_{hi}x_{hj}\left(\frac{y_{hi}}{x_{hi}} - \frac{y_{hj}}{x_{hj}}\right)^2$$

Where く$_{hij} = \sum_{s_h\ni i,j} d_{hsi}\, d_{hsj}\, p(s_h) - \sum_{s_h\ni i} d_{hsi}\, p(s_h) - \sum_{s_h\ni j} d_{hsj}\, p(s_h) + 1$

An unbiased estimator of $MSE(\widehat{Y_{lrc}})$ can be written as

$$\widehat{MSE}(\widehat{Y_{lrc}}) = -\sum_{h=1}^{L}\sum_{i<j=1}^{N_h}\sum_{j=1}^{N_h} d_{hij}\,(s_h)x_{hi}x_{hj}\left(\frac{y_{hi}}{x_{hi}} - \frac{y_{hj}}{x_{hj}}\right)^2$$

Where $d_{hij}\,(s_h) = 0$ if i and j do not belong to s_h or s_h does not contain i and j and $d_{hij}\,(s_h) = \frac{d_{hij}}{\pi_{hij}}$ if i and j belong to $s_h$or s_h contains i and j and π_{hij} is the inclusion probability of a pair of units i *and* j *and is given by* $\pi_{hij} \quad = \frac{n_h(n_h-1)}{N_h(N_h-1)}, h = 1,2, \ldots, L.$

9. MEAN SQUARE ERROR OF PRODUCT ESTIMATOR

The product estimator of the population total Y is given by

$$\widehat{Y_P} = N \frac{\bar{y}\bar{x}}{\bar{X}}$$

Where $\bar{y}$ and $\bar{x}$ are the sample means of y and x based on a sample of size n drawn from a population of size N by SRSWOR and $\bar{X}$ is the population mean of x.

Now $\widehat{Y_P}$ can be written as $\widehat{Y_P} = \sum_{i=1}^{N} d_{si}\, y_i$

Where $d_{si} = 0$ if i does not belong to s or s does not contain i and

$$d_{si} = \frac{N^2}{n} \frac{\bar{x}}{X} \text{ if } i \in s \text{ or } s \ni i$$

$$= \frac{N^2}{n^2} \frac{\sum_{i\in s} x_i}{X} = \frac{N^2}{n^2} \sum_{i\in s} p_i = \frac{N^2}{n^2} p_s, \text{ say, where } p_i = \frac{x_i}{X} \text{ and } p_s = \sum_{i\in s} p_i.$$

We may note that there does not exist a choice $y_i = cw_i$ for which $MSE(\widehat{Y_P}) = 0$. Thus the method of Rao (1979) fails here to derive an exact expression for the mean square error of $\widehat{Y_P}$.

In the next section we suggest an alternative procedure of deriving an exact expression for the variance of a homogeneous linear unbiased estimator of the population total in which one need not necessarily search for a choice $y_{i} = cw_i$ for which the variance of the estimator vanishes and there is no need of calculation of d_{ij} and $d_{ij}\,(s)$ as is required in the method suggested by Rao(1979).

10. AN ALTERNATIVE PROCEDURE OF DERIVING THE VARIANCE OF A HOMOGENEOUS LINEAR UNBIASED ESTIMATOR OF THE POPULATION TOTAL

Let $t = \sum_{i=1}^{N} d_{si}\, y_i$ be a homogeneous linear unbiased estimator of the population total Y where $d_{si} = 0$ if i does not belong to s or s does not contain i and $\sum_{s\ni i} d_{si}\, p(s) = 1\ \forall i.$

The variance of t is given by

$$Var(t) = E_p(t - Y)^2 = E_p(t^2) - Y^2$$

$$= E_p(\sum_{i=1}^{N} d_{si}\, y_i)^2 - Y^2$$

$$= E_p\left[\sum_{i=1}^{N} d_{si}^2 y_i^2 + \sum_{i=1}^{N} \sum_{j\neq i=1}^{N} d_{si} d_{sj}\, y_i y_j\right] - Y^2$$

$$= \sum_{s\in S}\left(\sum_{i=1}^{N} d_{si}^2 y_i^2\right) p(s) + \sum_{s\in S}\left(\sum_{i=1}^{N} \sum_{j\neq i=1}^{N} d_{si} d_{sj}\; y_i y_j\right) p(s) - Y^2$$

Where S is the collection of all possible samples s

$$= \sum_{s\in S}(\sum_{i\in s} d_{si}^2\, y_i^2)\, p(s) + \sum_{s\in S} \sum_{i\neq j\in s} \sum_{j\in s} d_{Ci}\, d_{sj}\, y_i y_j\, p(s) - Y^2$$

$$= \sum_{i=1}^{N} y_i^2\, \left[\sum_{s\ni i} d_{si}^2 p(s) - 1\right] + \sum_{i=1}^{N} \sum_{j\neq i=1}^{N} y_i\, y_j \left[\sum_{s\ni i,j} d_{si}\, d_{sj}\, p(s) - 1\right]$$

$= \sum_{i=1}^{N}]_i^2 \left[E_p(d_{si}^2) - 1\right] + \sum_{i=1}^{N}\sum_{j\neq i=1}^{N} y_i y_D \left[E_p\left(d_{si} d_{sj}\right) - 1\right].$

An unbiased estimator of $Var(t)$ is given by

$\widehat{Var}(t) = \sum_{i\in s} y_i^2 \left[d_{si}^2 - f_i(s)\right] + \sum_{i\in s}\sum_{j\neq i\in s} y_i\, y_j \left[d_{si} d_{sj} - f_{ij}(s)\right]$

Where $f_i(s)$ and $f_{ij}(s)$ are such that

$\sum_{s\ni i} f_i(s) p(s) = 1$ and $\sum_{s\ni i,j} f_{ij}(s)\, p(s) = 1.$

The possible choices of $f_i(s)$ and $f_{ij}(s)$ are

$f_i(s) = \frac{1}{\pi_i}, f_i(s) = \frac{1}{M_1 p(s)}$

And $f_{ij}(s) = \frac{1}{\pi_{ij}}, f_{ij}(s) = \frac{1}{M_2 p(s)}$

Where π_i and π_{ij} denote respectively the first order and second order inclusion probabilities and $M_r = \binom{N-r}{n-r}$, r=1, 2.Type equation here.

This procedure is applied to find the variances of the Horvitz-Thompson (1952) estimator, Hansen-Hurwitz (1943) estimator, Murthy's (1957) unordered estimator, ratio estimator based on Lahiri (1951), Midzuno (1952) and Sen's (1953) sampling scheme and Hartley – Ross (1954) unbiased ratio type estimator based on SRSWOR sampling scheme and also to find an exact expression for an unbiased estimator of the variance in each case.

11. VARIANCE OF HORVITZ - THOMPSON (1952) ESTIMATOR

The Horvitz-Thompson (1952) estimator is

$\widehat{Y_{HTE}} = \sum_{i\in s} \frac{y_i}{\pi_i}.$

Thus $\widehat{Y_{HTE}}$ can be written as $\widehat{Y_{HTE}} = \sum_{i=1}^{N} d_{si}\, y_i$ where d_{si} is zero if i does not belong to s or s does not contain i and $d_{si} = \frac{1}{\pi_i}$ if $i \in s$ or $s \ni i$.

The variance of $\widehat{Y_{HTE}}$ is

$Var\left(\widehat{Y_{HTE}}\right) = \sum_{i=1}^{N} y_i^2 \left[\sum_{s\ni i} d_{si}^2 p(s) - 1\right] + \sum_{i=1}^{N}\sum_{j\neq i=1}^{N} y_i\, y_j \left[\sum_{s\ni i,j} d_{si}\, d_{sj}\, p(s) - 1\right]$

$= \sum_{i=1}^{N} y_i^2 \left(\frac{1}{\pi_i} - 1\right) + \sum_{i=1}^{N}\sum_{j\neq i=1}^{N} y_i\, y_j \left[\frac{\pi_{ij}}{\pi_i \pi_j} - 1\right].$

An unbiased estimator of $Var\left(\widehat{Y_{HTE}}\right)$ is

$\widehat{Var}\left(\widehat{Y_{HTE}}\right) = \sum_{i\in s} y_i^2 \left(d_{si}^2 - \frac{1}{\pi_i}\right) + \sum_{i\in s}\sum_{j\neq i\in s} y_i\, y_j \left(d_{si} d_{sj} - \frac{1}{\pi_{ij}}\right)$

$= \sum_{i\in s} y_i^2 \left(\frac{1-\pi_i}{\pi_i^2}\right) + \sum_{i\in s}\sum_{j\neq i\in s} y_i\, y_j \left(\frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij} \pi_i \pi_j}\right).$

This expression is due to Horvitz-Thompson (1952).

Now to derive the other expressions for the variance of the Horvitz-Thompson(1952) estimator available in the literature, let us consider the following results.

Theorem 1: $\sum_{i=1}^{N} a_{ii}\, x_i^2 + \sum_{i=1}^{N}\sum_{j\neq i=1}^{N} a_{ij}\, x_i x_j = \sum_{i=1}^{N} \frac{x_i^2}{\alpha_i} A_i - \sum_{i<j=1}^{N}\sum_{j=1}^{N} a_{ij} \quad \alpha_i \quad \alpha_j \left(\frac{x_i}{\alpha_i} - \frac{x_j}{\alpha_j}\right)^2$ where $A_i = \sum_{j=1}^{N} a_{ij}\, \alpha_j$.

Proof: $\sum_{i<j=1}^{N}\sum_{j=1}^{N} a_{ij}\, \alpha_i \alpha_j \left(\frac{x_i}{\alpha_i} - \frac{x_j}{\alpha_j}\right)^2$

$= \frac{1}{2}\sum_{i\neq j=1}^{N}\sum_{j=1}^{N} a_{ij}\, \alpha_i \alpha_j \left[\frac{x_i^2}{\alpha_i^2} + \frac{x_j^2}{\alpha_j^2} - \frac{2x_i x_j}{\alpha_i \alpha_j}\right]$

$= \frac{1}{2}\left[2\sum_{i=1}^{N}\sum_{j\neq i=1}^{N} a_{ij}\, \frac{x_i^2}{\alpha_i}\alpha_j - 2\sum_{i\neq j=1}^{N}\sum_{j=1}^{N} a_{ij}\, x_i x_j\right]$

$= \sum_{i=1}^{N} \frac{x_i^2}{\alpha_i}\sum_{j\neq i=1}^{N} a_{ij}\, \alpha_j - \sum_{i\neq j=1}^{N}\sum_{j=1}^{N} a_{ij}\, x_i x_j$

$= \sum_{i=1}^{N} \frac{x_i^2}{\alpha_i}\sum_{j=1}^{N} a_{ij}\, \alpha_j - \sum_{i=1}^{N} a_{ii}\, x_i^2 - \sum_{i\neq j=1}^{N}\sum_{j=1}^{N} a_{ij}\, x_i x_j$

$\Rightarrow \sum_{i=1}^{N} a_{ii}\, x_i^2 + \sum_{i=1}^{N}\sum_{j\neq i=1}^{N} a_{ij}\, x_i x_j = \sum_{i=1}^{N} \frac{x_i^2}{\alpha_i} A_i - \sum_{i<j=1}^{N}\sum_{j=1}^{N} a_{ij}\, \alpha_i \alpha_j \left(\frac{x_i}{\alpha_i} - \frac{x_j}{\alpha_j}\right)^2$

Where $A_i = \sum_{j=1}^{N} a_{ij}\, \alpha_j$. That completes the proof.

Theorem 2: If $\sum_{i=1}^{N} \alpha_i = 1$, $\sum_{i<j=1}^{N}\sum_{j=1}^{N} \alpha_i\, \alpha_j \left(\frac{x_i}{\alpha_i} - \frac{x_j}{\alpha_j}\right)^2 = \sum_{i=1}^{N} \alpha_i \left(\frac{x_i}{\alpha_i} - X\right)^2$

Where $X = \sum_{i=1}^{N} x_i$.

Proof: Putting $a_{ij} = 1$ for all i and j in Theorem1, we get

$\sum_{i<j=1}^{N}\sum_{j=1}^{N} \alpha_i\, \alpha_j \left(\frac{x_i}{\alpha_i} - \frac{x_j}{\alpha_j}\right)^2 = \sum_{i=1}^{N} \frac{x_i^2}{\alpha_i} - \sum_{i=1} x_i^2 - \sum_{i\neq j=1}^{N}\sum_{j=1}^{N} x_i\, x_j$

$= \sum_{i=1}^{N} \frac{x_i^2}{\alpha_i} - X^2$ where $X = \sum_{i=1}^{N} x_i$

$= \sum_{i=1}^{N} \alpha_i \left(\frac{x_i}{\alpha_i} - X\right)^2$. That completes the proof.

Theorem 3: If either $A_i = 0$ for every $= 1,2,\ldots,N$ or $\sum_{i=1}^{N} \frac{x_i^2}{\alpha_i} A_i = 0$, then

$\sum_{i=1}^{N} a_{ii}\, x_i^2 + \sum_{i\neq j=1}^{N}\sum_{j=1}^{N} a_{ij}\, x_i x_j = -\sum_{i<j=1}^{N}\sum_{j=1}^{N} a_{ij}\, \alpha_i \alpha_j \left(\frac{x_i}{\alpha_i} - \frac{x_j}{\alpha_j}\right)^2$.

Proof: Follows from Theorem 1 by putting $A \quad = 0$ for every $= 1,2,\ldots,N$.

Now by virtue of Theorem 1, the variance of the Horvitz-Thompson(1952) estimator can be written as

$Var\left(\widehat{Y_{HTE}}\right) = \sum_{i=1}^{N} \frac{y_i^2}{\pi_i}\sum_{j=1}^{N} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j}\pi_j - \sum_{i<j=1}^{N}\sum_{j=1} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j}\pi_i \pi_j \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2$

By putting $x_i = y_i, \alpha_i = \pi_i$ and $a_{ij} = \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j}$ in Theorem 1 $= \sum_{i=1}^{N} \frac{y_i^2}{\pi_i}\sum_{j=1}^{N} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i} + \sum_{i<j=1}^{N}\sum_{j=1}^{N}\left(\pi_i \pi_j - \pi_{ij}\right)\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2$

$= \sum_{i=1}^{N} \frac{y_i^2}{\pi_i} \beta_i + \sum_{i<j=1}^{N} \sum_{j=1}^{N} (\pi_i \pi_j - \pi_{ij}) \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2$, say,

Where $\beta_i = \sum_{j=1}^{N} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i} = 1 + \frac{1}{\pi_i} \sum_{j \neq i=1}^{N} \pi_{ij} - \sum_{j=1}^{N} \pi_j$.

Now if we denote by $\gamma(s)$ the number of distinct units in s and if $\gamma(s)$ is constant for every sample s with $p(s) > 0$, then $\beta_i = 0$ for every $i = 1,2, \dots, N$

Because $\beta_i = 1 + \frac{1}{\pi_i} \sum_{j \neq i=1}^{N} \pi_{ij} - \sum_{j=1}^{N} \pi_j = 1 + \frac{(n-1)\pi_i}{\pi_i} - n$

$= 1 + (n-1) - n = 0$ as for a fixed effective sample size(n) design we have $\sum_{j \neq i=1}^{N} \pi_{ij} = (n-1)\pi_i$ and $\sum_{j=1}^{N} \pi_j = n$.

Thus for a fixed effective sample size(n) design, we have

$$Var(\widehat{Y_{HTE}}) = \sum_{i<j=1}^{N} \sum_{j=1}^{N} (\pi_i \pi_j - \pi_{ij}) \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2$$

And an unbiased estimator of the variance of $\widehat{Y_{HTE}}$ is given by

$$\widehat{Var}(\widehat{Y_{HTE}}) = \sum_{i<j \in s} \sum_{j \in s} \frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}} \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2$$

This expression is due to Yates and Grundy (1953).

Now if $\gamma(s)$ is not a constant for all s with $p(s) > 0$, then a third expression for the variance of the Horvitz-Thompson (1952) estimator is given by

$$Var(\widehat{Y_{HTE}}) = \sum_{i=1}^{N} \frac{y_i^2}{\pi_i} \beta_i + \sum_{i<j=1}^{N} \sum_{j=1}^{N} (\pi_i \pi_j - \pi_{ij}) \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2$$

And an unbiased estimator of the variance of $\widehat{Y_{HTE}}$ is given by

$$\widehat{Var}(\widehat{Y_{HT\square}}) = \sum_{i \in s} \frac{y_i^2}{\pi_i} \frac{\beta_i}{\pi_i} + \sum_{i<j \in s} \sum_{j \in s} \frac{\pi_i \pi_j - \pi_{ij}}{\pi_{ij}} \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2$$

This expression is due to Chaudhuri(2000). Chaudhuri and Pal (2002),Pal(2002) and Chaudhuri and Stenger(2005) claimed that the above expression of the variance estimator is uniformly non-negative if the following two conditions hold simultaneously

$\pi_i \pi_j \geq \pi_{ij}$ for all i and $j, i \neq j$

And $\beta_i > 0$ for all i.

But the above two conditions cannot hold simultaneously as is quite evident from the following result.

Theorem 4: If $\pi_i \pi_j \geq \pi_{ij}$ for all i and $j, i \neq j$, then $\beta_i < 0$ for all i.

Proof: $\beta_i = 1 + \frac{1}{\pi_i} \sum_{j \neq i=1}^{N} \pi_{ij} - \sum_{j=1}^{N} \pi_j$

$= \frac{\pi_i + \sum_{j \neq i=1}^{N} \pi_{ij} - \pi_i \sum_{j=1}^{N} \pi_j}{\pi_i}$

$= \frac{\sum_{j=1}^{N} \pi_{ij} - \sum_{j=1}^{N} \pi_i \pi_j}{\pi_i}$

$= \frac{1}{\pi_i}\left[\sum_{j=1}^{N}\left(\pi_{ij} - \pi_i\pi_j\right)\right] < 0$ for all i.

That completes the proof.

12. VARIANCE OF HANSEN – HURWITZ (1943) ESTIMATOR

The Hansen-Hurwitz (1943) estimator is

$$\widehat{Y_{HHE}} = \sum_{i=1}^{N} \frac{y_i f_i}{n p_i}$$

Where p_i is the normed size-measure of the ith unit and f_i is the frequency with which the ith unit occurs in the sample, $\sim = 1,2,\dots,N$.

Thus $\widehat{\langle_{HHE}}$ can be written as $\widehat{Y_{HHE}} = \sum_{i=1}^{N} d_{si}\, y_i$ where d_{si} is 0 if i does not belong to s or s does not contain i and $d_{si} = \frac{f_i}{np_i}$ if $i \in s$ or $s \ni i$.

The variance of the Hansen-Hurwitz (1943) estimator is

$$Var\left(\widehat{Y_{HHE}}\right) = \sum_{i=1}^{N} y_i^2 \left[E_p(d_{si}^2) - 1\right] + \sum_{i\neq j=1}^{N}\sum_{j=1}^{N} y_i y_j \left[E_p\left(d_{si} d_{sj}\right) - 1\right]$$

$$= \sum_{i=1}^{N} y_i^2 \left[E_p\left(\frac{f_i^2}{n^2 p_i^2}\right) - 1\right] + \sum_{i\neq j=1}^{N}\sum_{j=1}^{N} y_i y_j \left[E_p\left(\frac{f_i f_j}{n^2 p_i p_j}\right) - 1\right]$$

$$\sum_{i=1}^{N} y_i^2 \left[\frac{n(n-1)p_i^2}{n^2 p_i^2} + \frac{np_i}{n^2 p_i^2} - 1\right] + \sum_{i\neq j=1}^{N}\sum_{j=1}^{N} y_i y_j \left[\frac{n(n-1)p_i p_j}{n^2 p_i p_j} - 1\right]$$

$$= \sum_{i=1}^{N} y_i^2 \left[\frac{n-1}{n} + \frac{1}{np_i} - 1\right] + \sum_{i\neq j=1}^{N}\sum_{j=1}^{N} y_i y_j \left[\frac{n-1}{n} - 1\right]$$

$$= \frac{1}{n}\sum_{i=1}^{N} \frac{y_i^2}{p_i} - \frac{1}{n} Y^2$$

$$= \frac{1}{n}\sum_{i=1}^{N} p_i \left(\frac{y_i}{p_i} - Y\right)^2.$$

An unbiased estimator of $Var\left(\widehat{Y_{HHE}}\right)$ is given by

$$\widehat{Var}\left(\widehat{Y_{HH夏}}\right) = \sum_{\_=1}^{N} y_i^2 \left[d_{si}^2 - a_i(s)\right] + \sum_{i\neq j=1}^{N}\sum_{j=1}^{N} y_i y_j \left[d_{si} d_{sj} - a_{ij}(s)\right]$$

Where $a_i(s)$ and $a_{ij}(s)$ are such that $E_p[a_i(s)] = 1$ and $E_p\left[a_{ij}(s)\right] = 1$.

$$= \sum_{i=1}^{N} y_i^2 \left[\frac{f_i^2}{n^2 p_i^2} - \frac{f_i(f_i - 1)}{n(n-1)p_i^2}\right] + \sum_{i\neq j=1}^{N}\sum_{j=1}^{N} y_i y_j \left[\frac{f_i f_j}{n^2 p_i p_j} - \frac{f_i f_j}{n(n-1)p_i p_j}\right]$$

On Noting that $E_p[f_i(f_i - 1)] = n(n-1)p_i^2$ and $E_p\left[f_i f_j\right] = n(n-1)p_i p_j$

$$= \sum_{i=1}^{N} \frac{y_i^2 f_i}{n(n-1)p_i^2} - \sum_{i=1}^{N} \frac{y_i^2 f_i^2}{n^2(n-1)p_i^2} - \sum_{i\neq j=1}^{N}\sum_{j=1}^{N} \frac{y_i y_j f_i f_j}{n^2(n-1)p_i p_j}$$

$$= \frac{1}{n(n-1)}\left[\sum_{i=1}^{N} \frac{y_i^2 f_i}{p_i^2} - n\left(\sum_{i=1}^{N} \frac{y_i f_i}{np_i}\right)^2\right] = \frac{1}{n(n-1)}\sum_{i=1}^{N} f_i \left(\frac{y_i}{p_i} - \widehat{Y_{HHE}}\right)^2$$

Which matches with the traditional unbiased estimator of $Var(\widehat{,_{HHE}})$.

13. VARIANCE OF MURTHY'S (1957) UNORDERED ESTIMATOR

The Murthy's (1957) unordered estimator is

$$t_M = \sum_{i\in s} \frac{y_i p(s/i)}{p(s)}$$

Where $p(s/i)$ is the conditional probability of selecting a sample s given that the i th unit is selected on the first draw and $p(s)$ is the probability of selecting the unordered sample s.

Thus the Murthy's (1957) unordered estimator can be written as

$$t_M = \sum_{i=1}^{N} d_{si}\, y_i$$

Where $d_{si} = 0$ if s does not contain i or i does not belong to$s$ and $d_{si} = \frac{p(s/i)}{p(s)}$ if $i \in s$ or $s \ni i$.

The variance of the Murthy's (1957) unordered estimator is

$$Var(t_M) = \sum_{i=1}^{N} y_i^2 [\sum_{s\ni i} d_{si}^2\, p(s) - 1] + \sum_{i\neq j=1}^{N} \sum_{j=1}^{N} y_i y_j \left[\sum_{s\ni i,j} d_{si} d_{sj} p(s) - 1\right]$$

$$= \sum_{i=1}^{N} y_i^2 \left[\sum_{s\ni i} \frac{p^2(s/i)}{p(s)} - 1\right] + \sum_{i\neq j=1}^{N} \sum_{j=1}^{N} y_i y_j \left[\sum_{s\ni i,j} \frac{p(s/i)p(s/j)}{p(s)} - 1\right]$$

$$= \sum_{i=1}^{N} \frac{y_i^2}{p_i} A_i - \sum_{i<j=1}^{N} \sum_{j=1}^{N} \left[\sum_{s\ni i,j} \frac{p(s/i)p(s/j)}{p(s)} - 1\right] p_i p_j \left(\frac{y_i}{p_i} - \frac{y_j}{p_j}\right)^2$$

On putting $x_i = y_i, \alpha_i = p_i$ and $A_i = \sum_{j=1}^{N} a_{ij}\alpha_j = \sum_{j=1}^{N} \left[\sum_{s\ni i,j} \frac{p(s/i)p(s/j)}{p(s)} - 1\right] p_j$ in Theorem1.

Now $\sum_{i=1}^{N} \frac{y_i^2}{p_i} A_i = \sum_{i=1}^{N} \frac{y_i^2}{p_i} \sum_{j=1}^{N} \sum_{s\ni i,j} \frac{p(s/i)p(s/j)}{p(s)} p_j - \sum_{i=1}^{N} \frac{y_i^2}{p_i}$

$= \sum_{s\in S} \sum_{i,j\in s} \frac{y_i^2}{p_i} p_j \frac{p\left(\frac{s}{i}\right)p\left(\frac{s}{j}\right)}{p(s)} - \sum_{i=1}^{N} \frac{y_i^2}{p_i} = \sum_{s\in S} \sum_{i\in s} \frac{y_i^2}{p_i} \sum_{j\in s} \frac{p_j p(s/j)}{p(s)} p(s/i) - \sum_{i=1}^{N} \frac{y_i^2}{p_i} = \sum_{s\in S} \sum_{i\in s} \frac{y_i^2}{p_i} p(s/i) - \sum_{i=1}^{N} \frac{y_i^2}{p_i} = \sum_{i=1}^{N} \frac{y_i^2}{p_i} \sum_{s\ni i} p(s/i) - \sum_{i=1}^{N} \frac{y_i^2}{p_i}$

$= \sum_{i=1}^{N} \frac{y_i^2}{p_i} - \sum_{i=1}^{N} \frac{y_i^2}{p_i} = 0.$

Hence by virtue of Theorem 3, the variance of Murthy's (1957) unordered estimator can be written as

$$Var(t_M) = \sum_{i<j=1}^{N} \sum_{j=1}^{N} \left[1 - \sum_{s\ni i,j} \frac{p(s/i)p(s/j)}{p(s)}\right] p_i p_j \left(\frac{y_i}{p_i} - \frac{y_j}{p_{\square}}\right)^2$$

An unbiased estimator of $Va\square(t_M)$ is given by

$$\widehat{Var}(t_M) = \sum_{i<j\in s} \sum_{j\in s} \left[\frac{p(s/ij)}{p(s)} - \frac{p(s/i)p(s/j)}{p^2(s)}\right] p_i p_j \left(\frac{y_i}{p_i} - \frac{y_j}{p_j}\right)^2$$

On noting that $\sum_{s\ni i,j} p(s/ij) = 1$ where $p(s/ij)$ is the conditional probability of selecting the sample s given that the units i and j are selected on the first two draws. After some simplifications, we have

$$\widehat{Var}(t_M) = \frac{1}{\{p(s)\}^2} \sum_{i<j\in s} \sum_{j\in s} [p(s/ij) - p(s/i)p(s/j)] p_i p_j \left(\frac{y_i}{\square_i} - \frac{y_j}{p_j}\right)^2$$

Which matches with the traditional unbiased estimator of the variance of Murthy's,(1957) unordered estimator.

14. VARIANCE OF RATIO ESTIMATOR BASED ON LAHIRI (1951), MIDZUNO (1952) AND SEN'S (1953) SAMPLING SCHEME

The ratio estimator of the population total total Y is given by

$$t_R = \frac{\bar{y}}{\bar{x}}X = \frac{\sum_{i\in s} y_i}{\sum_{i\in s} x_i}X = \frac{\sum_{i\in s} y_i}{\sum_{i\in s} p_i} = \frac{\sum_{i\in s} y_i}{p_s} \text{ where } p_s = \sum_{i\in s} p_i.$$

The ratio estimator can be written as

$$t_R = \sum_{i=1}^{N} d_{si}\, y_i$$

Where d_{si} is 0 if i does not belong to s or s does not contain i and

$$d_{si} = \frac{1}{p_s} \text{ if } i \in s \text{ or } s \ni i.$$

The probability of selecting a sample s according to Lahiri (1951), Midzuno (1952) and Sen's (1953) sampling scheme is given by

$$p(s) = \frac{p_s}{M_1} \text{ where } M_1 = \binom{N-1}{n-1}.$$

The variance of the ratio estimator based on Lahiri(1951), Midzuno(1952) and Sen's(1953) sampling scheme is given by

$$Var(t_R) = \sum_{i=1}^{N} y_i^2 \left[\sum_{s\ni i} d_{si}^2\, p(s) - 1\right] + \sum_{i\neq j=1}^{N}\sum_{j=1}^{N} y_{i} y_j \left[\sum_{s\ni i,j} d_{si} d_{sj}\, p(s) - 1\right]$$

$$= \sum_{i=1}^{N} y_i^2 \left[\sum_{s\ni i} \frac{1}{p_s}\frac{1}{M_1} - 1\right] + \sum_{i\neq j=1}^{N}\sum_{j=1}^{N} y_i y_j \left[\sum_{s\ni i,j} \frac{1}{p_s}\frac{1}{M_1} - 1\right]$$

$$= \sum_{i=1}^{N} \frac{y_i^2}{p_i} A_i - \sum_{i<j=1}^{N}\sum_{j=1}^{N}\left[\sum_{s\ni i,j} \frac{1}{p_s}\frac{1}{M_1} - 1\right] p_i p_j \left(\frac{y_i}{p_i} - \frac{y_j}{p_j}\right)^2$$

On putting $x_i = y_i, \alpha_i = p_i$ and $A_i = \sum_{j=1}^{N} a_{ij}\alpha_j = \sum_{j=1}^{N}\left[\sum_{s\ni i,j} \frac{1}{p_s}\frac{1}{M_1} - 1\right] p_j$ in Theorem1

Now let us consider the quantity

$$\sum_{i=1}^{N} \frac{y_i^2}{p_i} A_i = \sum_{i=1}^{N} \frac{y_i^2}{p_i} \sum_{j=1}^{N}\left[\sum_{s\ni i,j} \frac{1}{p_1}\frac{1}{M_1} - 1\right] p_j$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{s\ni i,j} \frac{1}{p_s}\frac{1}{M_1}\frac{y_i^2}{p_i} p_j - \sum_{i=1}^{N}\frac{y_i^2}{p_i}$$

$$= E_p\left[\sum_{i\in s}\sum_{j\in s} \frac{1}{p_s^2}\frac{y_i^2}{p_i} p_j\right] - \sum_{i=1}^{N}\frac{y_i^2}{p_i}$$

$$= E_p\left[\sum_{i\in s} \frac{y_i^2}{p_i}\frac{1}{p_s}\right] - \sum_{-=1}^{N}\frac{y_i^2}{p_i}$$

$$= \sum_{s\in S}\sum_{i\in s} \frac{y_i^2}{p_i}\frac{1}{p_s} p(s) - \sum_{i=1}^{N}\frac{y_i^2}{p_i}$$

Where S is the collection of all possible samples s

$$= \sum_{s\in S}\sum_{i\in s} \frac{y_i^2}{p_i}\frac{1}{p_s}\frac{p_s}{M_1} - \sum_{i=1}^{N}\frac{y_i^2}{p_i}$$

$$= \sum_{i=1}^{N} \frac{y_i^2}{p_i}\sum_{s\ni i}\frac{1}{M_1} - \sum_{i=1}^{N}\frac{y_i^2}{p_i}$$

$$= \sum_{i=1}^{N} \frac{y_i^2}{p_i} - \sum_{i=1}^{N} \frac{y_i^2}{p_i} = 0$$

Hence by Theorem 3 the variance of the ratio estimator can be written as

$$Var(t_R) = \sum_{i<j=1}^{N} \sum_{j=1}^{N} \left[1 - \sum_{s \ni i,j} \frac{1}{p_s} \frac{1}{M_1}\right] p_i p_j \left(\frac{y_i}{p_i} - \frac{y}{p_j}\right)^2$$

An unbiased estimator of the variance of t_R is given by

$$\widehat{Var}(t_R) = \sum_{i<j \in s} \sum_{j \in s} \left[\frac{1}{M_2 p(s)} - \frac{1}{p_s^2}\right] p_i p_j \left(\frac{y_i}{p_i} - \frac{y_j}{p_j}\right)^2$$

$$= \sum_{i<j \in s} \sum_{j \in s} \left[\frac{M_1}{M_2 p_s} - \frac{1}{77_s^2}\right] p_i p_j \left(\frac{y_i}{p_i} - \frac{y_j}{p_j}\right)^2$$

Where $M_2 = \binom{N-2}{n-2}$.

A sufficient condition for non-negativity of $\widehat{Var}(t_R)$ is

$$p_s > \frac{M_2}{M_1}$$

or$\frac{\sum_{i \in s} x_i}{X} > \frac{n-1}{N-1}$

or$\sum_{i \in s} x_i > \frac{n-1}{N-1}$暆

or$\underset{1 \le i \le n}{Min} x_i > \frac{n-1}{N-1} \frac{X}{n}$

As noted by Rao and Vijayan (1977).

15. VARIANCE OF HARTLEY –ROSS (1954) UNBIASED RATIO TYPE ESTIMATOR

The Hartley-Ross (1954) unbiased ratio type estimator of the population total Y is given by

$$\widehat{Y_{HR}} = \bar{r}X + \frac{n(N-1)}{n-1}(\bar{y} - \bar{r}\bar{x})$$

Where $\bar{r} = \frac{1}{n}\sum_{i=1}^{n} \frac{y_i}{x_i}$ is the mean of the ratios and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$,$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ are the sample means of the study variable y and the auxiliary variable x respectively. With the same notations as above Chaudhuri and Stenger (2005) have provided an incorrect expression for the Hartley-Ross (1954) unbiased ratio type estimator for the population total as

$$\widehat{Y_{HR}} = \bar{r}X + \frac{n(N-1)}{N(n-1)}(\bar{y} - \bar{r}\bar{x})$$

In which the coefficient of the second term involving $(\bar{y} - \bar{r}\bar{x})$ is appropriate for estimating the population mean $\bar{Y} = \frac{Y}{N}$ rather than Y in which case the first term should be replaced by $\bar{r}\bar{X}$.

The Hartley-Ross(1954) unbiased ratio type estimator can be written as

$$\widehat{Y_{HR}} = \sum_{i=1}^{N} d_{sh}\, y_i$$

Where $d_{si} = 0$ if i does not belong to s or s does not contain i and

$$d_{si} = \frac{1}{n}\frac{X}{x_i} + \frac{n(N-1)}{n-1}\left(\frac{1}{n} - \frac{1}{n}\frac{\bar{x}}{x_i}\right)$$

If $i \in s$ or $s \ni i$.

The variance of the Hartley-Ross (1954) unbiased ratio type estimator is given by

$$Var(\widehat{Y_{HR}}) = \sum_{i=1}^{N} y_i^2 \left[\sum_{s\ni i} d_{si}^2\, p(s) - 1\right] + \sum_{i\neq j=1}^{N} \sum_{j=1}^{N} y_i y_j \left[\sum_{s\ni i,j} d_{si} d_{sj}\, p(s) - 1\right]$$

$$= \sum_{i=1}^{N} \frac{y_i^2}{x_i} A_i - \sum_{i<j=1}^{N} \sum_{j=1}^{N} a_{ij}\, x_i x_j \left(\frac{y_i}{x_i} - \frac{y_j}{x_j}\right)^2$$

On putting $x_i = y_i$, $\alpha_i = x_i$ and $A_i = \sum_{j=1}^{N} a_{ij}\, x_j$ in Theorem 1 where

$$a_{ij} = \sum_{s\ni i,j} d_{si} d_{sj}\, p(s) - 1$$

Now consider the quantity

$$\sum_{i=1}^{N} \frac{y_i^2}{x_i} A_i = \sum_{i=1}^{N} \frac{y_i^2}{x_i} \sum_{j=1}^{N} a_{ij}\, x_j$$

$$= \sum_{i=1}^{N} \frac{y_i^2}{x_i} \sum_{j=1}^{N} \left[\sum_{s\ni i,j} d_{si} d_{sj}\, p(s) - 1\right] x_j$$

$$= \sum_{s\in S} \sum_{i\in s} \sum_{j\in s} d_{si} d_{sj}\, x_j \frac{y_i^2}{x_i} p(s) - \sum_{i=1}^{N} \frac{y_i^2}{x_i} X$$

Where S is the collection of all possible samples

$$= \sum_{s\in S} \sum_{i\in s} \frac{y_i^2}{x_i} d_{si}\, p(s) \sum_{j\in s} d_{sj}\, x_j - \sum_{j=1}^{N} \frac{y_i^2}{x_i} X$$

$$= \sum_{s\in S} \sum_{i\in s} \frac{y_i^2}{x_i} d_{si}\, p(s) X - \sum_{i=1}^{N} \frac{y_i^2}{x_i} X$$

Because of the calibration equation $\sum_{j\in s} d_{sj}\, x_j = X$

$$= \sum_{i=1}^{N} \frac{y_i^2}{x_i} X \sum_{s\ni i} d_{si}\, p(s) - \sum_{i=1}^{N} \frac{y_i^2}{x_i} X$$

$$= \sum_{i=1}^{N} \frac{y_i^2}{x_i} X - \sum_{i=1}^{N} \frac{y_i^2}{x_i} X = 0$$

Using the unbiasedness condition $\sum_{s\ni i} d_{si}\, p(s) = 1 \forall i = 1,2,\ldots,N$ as the Hartley-Ross (1954) estimator is known to be unbiased under SRSWOR.

Hence by Theorem3 the variance of the Hartley-Ross (1954) unbiased ratio estimator is given by

$$Var(\widehat{Y_{HR}}) = \sum_{i<j=1}^{N} \sum_{j=1}^{N} \left[1 - \sum_{s\ni i,j} d_{si} d_{sj}\, p(s)\right] x_i x_j \left(\frac{y_i}{x_i} - \frac{y_j}{x_j}\right)^2$$

An unbiased estimator of the variance of Hartley-Ross (1954) unbiased ratio type estimator is given by

$$\widehat{Var}(\widehat{Y_{HR}}) = \sum_{i<j\in s} \sum_{j\in s} \left[\frac{1}{\pi_{ij}} - d_{si} d_{sj}\right] x_i x_j \left(\frac{y_i}{x_i} - \frac{y_j}{x_j}\right)^2$$

Where $\pi_{ij} = \frac{n(n-1)}{N(N-1)}$ is the inclusion probability of a pair of units i and j under SRSWOR.

16. CONCLUSIONS

Thus with this newly suggested procedure it is possible to derive the variance of a homogeneous linear unbiased estimator $\hat{Y} = \sum_{i=1}^{N} d_{si}\, y_i$ of the population total Y based on a sampling design $p(s)$ under which it is unbiased and it is also possible to derive an unbiased estimator of the variance of the estimator based on the same sampling design(s). In this procedure there is no need to search for a choice $y_i = cw_i$for which $Var(\hat{Y}) = 0$ and also there is no need to calculate d_{ij} or $d_{ij}(s)$ as is required in Rao's (1979) procedure. That is the novelty of this newly suggested technique which has been illustrated in case of several well-known estimators of the population total based on varying probability sampling designs including the Hartley-Ross (1954) unbiased ratio type estimator based on SRSWOR sampling design for which neither an expression for the variance of the estimator nor an unbiased estimator of the variance is available in the literature on Survey Sampling.

17. REFERENCES

1. **A. K. Adhikary,** Small Area Estimation – A review, Sankhyikee, 7, pp.41-58, 2000.
2. **A. K. Adhikary**, Small Area Estimation – A review, Journal of the Indian Society for Probability and Statistics 9, pp.17-34, 2005.
3. **C. M. Cassel, C. E. S$\ddot{a}$rndal and J. H. wretman,** Some results on generalized difference estimation and generalized regression estimation for finite populations for finite populations, Biometrika, 63, pp. 615 - 620, 1976.
4. **A. Chaudhuri and A.K. Adhikary,** On Generalized Regression Estimation of Small Domain Totals–An Evaluation Study, Pakistan Journal of Statistics, 11, 3, pp. 173-189, 1995.
5. **A. Chaudhuri and A. K. Adhikary,** On Generalized Regression Predictors of Small Domain Totals – An Evaluation Study in Indian Statistical Institute, Calcutta, Frontiers in Probability and Statistics, pp.100-105, Narosa Publishing House, Edited by S. P. Mukherjee, S. K. Basu and B. K. Sinha, 1998.
6. **A. Chaudhuri,** Network and adaptive sampling with unequal probabilities, Calcutta Statistical Association Bulletin, pp. 237 – 253, 2000.
7. **A**. **Chaudhuri and S. Pal,** On certain alternative mean square errors in complex surveys, Journal of Statistical Planning and Inference, 104, 2, pp. 363 – 375, 2002.
8. **A. Chaudhuri and H. Stenger,** Survey Sampling – Theory and Methods, Second Edition, Chapman & Hall /CRC,Taylor & Francis Group, New York, 2005
9. **M. H. Hansen and W. N. Hurwitz**, On the theory of sampling from finite populations, The Annals of Mathematical Statistics, 14, pp. 333 - 362, 1943.
10. **H. O. Hartley and A. Ross,** Unbiased ratio estimates, Nature, 174, pp. 270-271, 1954.
11. **D. G. Horvitz and D. J. Thompson,** A generalization of sampling without replacement from a finite universe, Journal of the American Statistical Association, 47, pp. 663 – 685, 1952.
12. **D. B. Lahiri,** A method of sample selection providing unbiased ratio estimators, Bulletin of International Statistical Institute, 33, 2, pp. 133 – 140, 1951.
13. **H. Midzuno,** On the sampling system with probabilities proportionate to sum of sizes, Annals of the Institute of Statistical Mathematics, 3, pp. 99 -107, 1952.
14. **M. N. Murthy,** Ordered and Unordered estimators in sampling without replacement, Sankhy$\bar{a}$, 18, pp. 379 – 390, 1957.

15. **S. Pal,** Contributions to emerging techniques in survey sampling, Unpublished Ph.D. Thesis submitted to Indian Statistical Institute, Kolkata, 2002.

16. **J. N. K. Rao,** On deriving mean square errors and their non-negative unbiased estimators in finite population sampling, Journal of the Indian Statistical Association, 17, pp.125 – 136, 1979.

17. **J. N. K. Rao and K. Vijayan,** On estimating the variance in sampling with probability proportional to aggregate size, Journal of the American Statistical Association, 72, 359, pp. 579 – 584, 1977.

18. **C. E. Särndal,** On π-inverse weighting versus best linear weighting in probability sampling**,** Biometrika, 67, pp. 639 – 650, 1980.

19. **C. E. Särndal,** Implications of survey design for generalized regression estimation of linear functions, Journal of Statistical Planning and Inference, 7, pp. 155 – 170, 1982.

20. **A. R. Sen,** On the estimation of variance in sampling with varying probabilities, Journal of the Indian Society of Agricultural Statistics, 5, 2, pp**.** 119 – 127, 1953.

21. **F. Yates and P. M. Grundy,** Selection without replacement from within strata with probability proportional to size**,** Journal of the Royal Statistical Society**,** Series B, 15, pp. 253 – 261, 1953.