UNBIASED RATIO-TYPE ESTIMATION IN SIMPLE RANDOM SAMPLING FROM STRATIFIED POPULATIONS

By

I. P. David[1] and B. V. Sukhatme[2]

(Received : October, 1976)

1. Introduction

In simple random sampling from bivariate finite populations, an unbiased ratio-type estimator y' of the mean has been found by Hartley and Ross [3]. Goodman and Hartley [2] give the variance of y' for large or infinite populations and Robson [8] gives the exact formula in the case of finite populations.

When the stratum means of the auxiliary variate x are known, the application of y' in stratified simple random sampling is straightforward. When only the population mean of x is known, Nieto de Pascual [6] gives a "combined" unbiased ratio estimator of the mean of y, which however, is built from an equal probability with complete replacement sampling scheme with the same sample size per stratum and hence is not directly analogous to the Hartley-Ross estimator. When sampling is with unequal probability without replacement, Nanjamma, Murthy and Sethi [5] give a combined analogue of the Lahiri unbiased ratio estimator [4].

We present here, together with its variance for large or infinite populations, a direct analogue of the Hartley-Ross estimator, namely a combined unbiased ratio-type estimator in simple random sampling from a stratified population.

2. A Combined Unbiased Ratio-Type Estimator

Consider drawing a simple random sample of n_i out of the N_i pairs (x_{ij}, y_{ij}) in the *i-th* stratum, $i=1, 2, \ldots\ldots, L$.

Let $r_{ij}=y_{ij}/x_{ij}$, $N=N_1+\ldots\ldots+N_L$ and denote by

$\bar{x}_N$, $\bar{y}_N$, $\bar{r}_N$ the poulation means.

$\bar{x}_{N_i}$, $\bar{y}_{N_i}$, $\bar{r}_{N_i}$ the i-the stratum means.

---

1. Associate Professor of Statistics, University of the Philippines at Los Banos, Philippines.

2. Professor of Statistics, Iowa State University, Ames, Iowa, U.S.A.

and $\bar{x}_{n_i}$, $\bar{y}_{n_i}$, $\bar{r}_{n_i}$ the i-th stratum sample.

Further, let

$$\bar{r}_{st}=\Sigma w_i \bar{r}_{n_i}\ ,\ w_i=N_i/N$$

and define $\bar{x}_{st}$ and $\bar{y}_{st}$ similarly for x and y, respectively. We address ourselves to the problem of estimating $\bar{y}_N$ unbiasedly given the value of $\bar{x}_N$.

Consider the estimator

$$t=\bar{r}_{st}\bar{x}_N. \qquad ...(2.1)$$

Clearly, t is biased since

$$\text{Bias }(t)=E(t)-\bar{y}_N=\bar{r}_N\ \bar{x}_N-\sum_i w_i \bar{y}_{N_i}$$

$$=-\frac{1}{N}\left(\sum_i\sum_j r_{ij}x_{ij}-N\bar{r}_N\ \bar{x}_N\right)$$

$$=-\sigma_{rx} \qquad ...(2.9)$$

The natural step to follow at this point is to find an unbiased estimator for σ_{rx} so that t may be adjusted for bias. taking note not to bring in the argument the unknown stratum means of the auxiliary variate x.

We do this by writing

$$\sigma_{rx}=\frac{1}{N}\left\{\sum_i N_i\left(\bar{r}_{N_i}-\bar{r}_N\right)\left(\bar{x}_{N_i}-\bar{x}_N\right)\right.$$

$$\left.+\sum_i\sum_j\left(r_{ij}-\bar{r}_{N_i}\right)\left(x_{ij}-\bar{x}_{N_i}\right)\right\}$$

$$=\sum_i w_i\bar{r}_{N_i}\bar{x}_{N_i}-\bar{r}_N\ \bar{x}_N+\sum_i\frac{N_i-1}{N}S_{irx} \qquad ...(2.3)$$

where S_{irx} is the convariance between r and x in the i-th stratum, which is estimated unbiasedly by

$$s_{irx}=\frac{n_i}{n_i-1}\left(\bar{y}_{n_i}-\bar{r}_{n_i}\bar{x}_{n_i}\right) \qquad ...(2.4)$$

since

$$\bar{r}_{N_i}\bar{x}_{N_i}=E\left(\bar{r}_{n_i}\bar{x}_{n_i}\right)-\text{Cov}\left(\bar{r}_{n_i},\ \bar{x}_{n_i}\right)$$

it follow that

$$\bar{r}_{n_i}\bar{x}_{n_i}-\left(\frac{1}{n_i}-\frac{1}{N_i}\right)s_{irx} \qquad \ldots(2.5)$$

is an unbiased estimator of $\bar{r}_{N_i}\bar{x}_{N_i}$.

Similarly,

$$\bar{r}_{st}\bar{x}_{st}-\sum w_i^2\left(\frac{1}{n_i}-\frac{1}{N_i}\right)s_{irx} \qquad \ldots(2.6)$$

is an unbiased estimator of

$$\bar{r}_N\bar{x}_N=E\left(\bar{r}_{st}\,\bar{x}_{st}\right)-\text{Cov}\left(\bar{r}_{st},\,\bar{x}_{st}\right).$$

Hence from (3.4), (2.5) and 2.6 and after some manipulation, an unbiased estimator of (2.3) is given by

$$\hat{\sigma}_{rx}=\sum_i w_i\left\{\bar{r}_{n_i}\bar{x}_{n_i}+\left(w_i-1\right)\left(\frac{1}{n_i}-\frac{1}{N_i}\right)s_{irx}\right\}$$

$$+\sum\frac{N_i-1}{N}s_{irx}-\bar{r}_{st}\bar{x}_{st}$$

$$=\sum w_i\left[\bar{r}_{ni}\,\bar{x}_{ni}+\left\{\frac{N-1}{N}-\frac{(1-w_i)}{n_i}\right\}s_{irx}\right]-\bar{r}_{st}\bar{x}_{st}.$$

Using the relation

$$\bar{r}_{n_i}\bar{x}_{n_i}=\bar{y}_{n_i}-s_{irx}(n_i-1)/n_i,\ \hat{\sigma}_{rx}$$

can be written further as

$$\hat{\sigma}_{rx}=\bar{y}_{st}-\bar{r}_{st}\bar{x}_{st}+\sum_i w_i^2\left(\frac{1}{n_i}-\frac{1}{N_i}\right)s_{irx} \qquad \ldots(2.7)$$

Thus the sum of (2.7) and t in (2.1) gives

$$y'_c=\bar{y}_{st}+\bar{r}_{st}\left(\bar{x}_N-\bar{x}_{st}\right)+\sum w_i^2\left(\frac{1}{n_i}-\frac{1}{N_i}\right)s_{irx} \qquad \ldots(3.8)$$

which is an unbiased estimator of $\bar{y}_N$.

To the best of the authors knowledge, the derivation given above of y'_c was first formulated in [1], where the variance of y'_c is also given. Using other approaches, Ross [9] and Rao [7] also have found y'_c earlier. The former author likewise gave a somewhat complicated formula for Var (y'_c) in the special case where the starta are of equal size and simple random sample of equal size are drawn each stratum. A derivation of Var (y'_c) in simple random sampling from very large or infinite stratum sizes is given in the appendix,

EMPIRICAL COMPARISON WITH THE HARTLEY-ROSS ESTIMATOR

The "separate", Hartley-Ross estimator

$$y'_s=\sum w_i\left\{\bar{r}_{n_i}\bar{x}_{N_i}+\frac{(N_i-1)n_i}{N_i(n_i-1)}\left(\bar{y}_{n_i}-\bar{r}_{n_i}\bar{x}_{n_i}\right)\right]$$

has variance [2]

$$\text{Var }(y_s')=\sum\frac{w_i^2}{n_i}\left[\bar{r}^2_{N_i}\sigma^2_{ix}-2\bar{r}_{N_i}\sigma_{ixy}+\sigma^2_{iy}+\frac{1}{n_i-1}\left(\sigma^3_{ir}\sigma^2_{ix}+\sigma^2_{irx}\right)\right] \quad ...(3.1)$$

while

$$\begin{aligned}\text{Var }(y_c')=&\sum\frac{w^2_i}{n_i}\left(\sigma^2_{iy}-2\bar{r}_N\,\sigma_{ixy}+\bar{r}^2_N\,\sigma^2_{ix}\right)\\&+\sum\frac{w^4_i}{n^2_i(n_i-1)}\left(\sigma^2_{ir}\,\sigma^2_{ix}+\sigma^2_{irx}\right)\\&+\left(\sum\frac{w_i\sigma^2_{ir}}{n_i}\right)\left(\sum\frac{w^2_i\sigma^2_{ix}}{n_i}\right)\\&+\left(\sum\frac{w^2_i\sigma_{irx}}{n_i}\right)^2\end{aligned} \quad ...(3.2)$$

Judging from the complex forms of these variances, an attempt to compare analytically the precisions of these two estimators without imposing imparctical restrictions will entail extreme algebraic difficulculties. On the other hand, since Var (y_c') and Var (y_s') are exact and not "asymptotic" expansions as in the case of other ratio-type estimators based on ratios of sample means, it may not be advisable to use large-sample formulas, *i.e.*, terms of other n_i^{-1}, to compare variances. Thus we resort to numercal comparisons:

(*a*) First, we use the 1959 (X) and 1964 (Y) censuses of agriculture for the state of Iowa. The 99 counties of the state were grouped into four strata of sizes 21, 20, 30 and 28, each consisting of geographically contiguous and economically similar counties. The following nine pairs of variables with correlations ranging from ·55 to ·98 were chosen for the study :

(X_1, Y_1) = (1959 acreage under corn, 1964 acreage under corn)

(X_2, Y_2) = (1959 value of farm products, 1964 value of farm products)

(X_3, Y_3) = (1964 number of farms, 1964 area of farms)

(X_4, Y_4) = (1964 number of farms, 1964 number of full farm owners)

(X_5, Y_5) = (1959 tons of fertilizer used, 1964 tons of fertilizer used)

(X_6, Y_6) = (1959 number of farms, 1964 farm expenditure for oil and fuel)

(X_7, Y_7) = (1959 number of farms reporting cattle, 1964 number of cattle)

(X_8, Y_8) = (1964 acreage under soybean, 1964 soybean production)

(X_9, Y_9) = (1959 number of farms, 1964 non-farm income).

For the purposes of comparison we ignored the finite correction factors, used proportional allociation and considered three sample sizes, namely 20, 30 and 40 counties for the whole state.

(*b*) Second, we consider the estimation of population counts using as example the 1960 (X) and 1970 (Y) censuses of population of the Philippines. For statistical and other purposes, the Philippines is divided into 9 regions (excluding region 1—Metropolitan Manila) each consisting of a number of provinces with similar ethnic and economic characteristics. We look at the estimation of the population of each region, with provinces as strata and and towns within provinces as sampling units. (This numerical exercise has practical relevance in the Philippines wherein results of the latest population census serve as frame for designing the quarterly household surveys. These latter surveys have utilized three stage sampling schemes with towns, barrios or villages and housholds as first, second and third stage units. The provinces are treated as separate domains of study and an average of 5 sample towns are drawn from each. Hence ratio estimation at the town level is used to improve the efficiency of provincial, regional and county estimates.) Further, we consider equal allocation (3, 4 and 5) of sample towns per stratum and igrone finite correction factors. The correlations between the 1960 and 1970 town populations for the 58 different strata range from ·80 to ·99.

The actual relative efficiencies of y_c' over y_s'

$$\text{R.E.} = \frac{\text{Var }(y_s')}{\text{Var }(y_c')} \times 100$$

are given in Tables 1*a* and 1*b*. It is seen that for smallish samples, y_c' is generally work efficient than y_s'. With population counts whose distributions are markedly skewed, the former estimator can be considerably more efficient (Table 1*b*) although there are instances where the latter is slightly more efficient. With data from less skewed poplations like agriculture data, Var (y_c') is consistently smaller. The advanlage in efficiency of y_c' diminishes as the sample size increases.

To terms of order n_i^{-1}, the inequality Var $(y_s')<$Var (y_c') is more order true than not. On the other hand, the contribution of the second order terms to the variance is much bigger for y_s' as seen in Tables 2*a* and 2*b*. Moreover, there are some cases where the sum of the $0\ (n_i^{-2})$ terms is even greater than that of the $0\ (n_i^{-1})$ terms. In fact, the results in these tables pose the warning that the $0(n_i^{-2})$ terms in Var (y_s') and Var (y_c'), more especially in the former, should not be assumed negligible even for moderate sample sizes.

TABLE 1*a*

Relative efficiencies in percent, Iowa agriculture data

Total sample size	*Variables*								
	(X_1, Y_1)	(X_2, Y_2)	(X_3, Y_3)	(X_4, Y_4)	(X_5, Y_5)	(X_6, Y_6)	(X_7, Y_7)	(X_8, Y_8)	(X_9, Y_9)
20	121	142	126	116	126	127	118	143	122
30	110	128	115	108	118	119	109	124	113
40	105	120	111	105	114	115	104	115	106

TABLE 1*b*

Relative efficiencies in percent, Philippines population data

Sample size per stratum	*Region number*								
	2	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
3	105	187	98	163	92	115	124	187	293
4	96	167	95	143	89	110	117	164	248
5	92	156	94	132	88	108	113	152	225

TABLE 2*a*

Ratios of 6 (n_i^{-2}) to 0 (n_i^{-1}) terms in the Variance, Iowa agriculture data

Estimator	*Total sample size*	*Variables*								
		(X_1, Y_1)	(X_2, Y_2)	(X_3, Y_3)	(X_4, Y_4)	(X_5, Y_5)	(X_6, Y_6)	(X_7, Y_7)	(X_8, Y_8)	(X_9, Y_9)
y_s'	20	.47	.74	.40	.31	.29	.28	.36	1.10	.30
	30	.29	.46	.24	.18	.18	.17	.22	.67	.18
	40	.21	.33	.17	.14	.13	.12	.16	.49	.13
y_c'	20	.11	.15	.08	.07	.07	.07	.07	.31	.06
	30	.07	.10	.05	.04	.04	.04	.04	.21	.04
	40	.05	.07	.04	.03	.03	.03	.03	.15	.03

TABLE 2*b*

Ratios of 0 (n_i^{-2}) to 0 (n_i^{-1}) terms in the variance, Philippines population data

Estimator	*Sample size per stratum*	*Region Number*								
		2	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
y_s'	3	.45	1.37	.11	4.48	.13	.14	.25	.80	.80
	4	.27	.92	.08	2.99	.08	.09	.17	.53	.53
	5	.20	.68	.06	2.24	.06	.07	.12	.40	.40
y_c'	3	.05	.54	.01	2.05	.03	.01	.03	.12	.13
	4	.04	.39	.01	1.52	.03	.01	.02	.09	.09
	5	.03	.31	.01	1.21	.02	.01	.02	.07	.07

APPENDIX

Let σ_{ir}^2, σ_{ix}^2 and σ_{iy}^2 be variances and σ_{irx}, σ_{ixy} be covariances in the *i-th* stratum. Ignoring finite correction factors, we now show that the variance of y_c' can be written as in (3.2).

It is postible and less cumbersome perhaps to find the exact Var (y_c') through the application of symmetric means and polykays [10] similar to what has been done by Robson [8] for the Hartley-Ross estimator in non-stratified sampling. The resulting variance form however, will be difficult to use for purposes of comparison and actual computation. Hence we use straightforward algebraic methods to derive Var (y_c') and when the use of symmetric means seems unavoidable, the end results are transformed to the usual product moments of x, y and r. We may add also that, when finite correction factors cannot be ignored, Var (y_c') can be obtained directly from the following proof but such variance cannot be expressed in simple form.

We write y_c' as

$$y_c' = \sum_i w_i \left\{ \bar{y}_{n_i} + \bar{r}_{n_i}\bar{x}_N - \bar{r}_{n_i}\bar{x}_{st} + w_i \left(\frac{1}{n_i} - \frac{1}{N_i} \right) \frac{n_i}{n_i - 1} \left(\bar{y}_{n_i} - \bar{r}_{n_i}\bar{x}_{n_i} \right) \right\}$$

$$= \sum_i w_i \left(A_i \bar{y}_{n_i} - B_i \bar{r}_{n_i} \bar{x}_{n_i} + \bar{r}_{n_i} \bar{x}_N \right) - \sum_{i \neq j} w_i w_j \bar{r}_{n_i} \bar{x}_{n_j} \quad \ldots(A.1)$$

where as $N_i \to \infty$,

$$A_i = 1 + \frac{w_i(N_i - n_i)}{N_i(n_i - 1)} \to 1 + \frac{w_i}{n_i - 1}$$

and

$$B_i = w_i \left\{ 1 + \frac{(N_i - n_i)}{N_i(n_i - 1)} \right\} \to \frac{n_i w_i}{n_i - 1}.$$

For non-negative integers a, b and c, let

$$\mu_{abc_i} = \frac{1}{N_i} \sum_j^{N_i} \left(x_{ij} - \bar{x}_{N_i} \right)^a \left(y_{ij} - \bar{y}_{N_i} \right)^b \left(r_{ij} - \bar{r}_{N_i} \right)^c$$

and following established notations [8, 10], we denote the trivariate symmetric function

$$\frac{(1)}{(N_i)_k} \sum_{j_1 \neq \ldots \neq j_k}^{N_i} \left(x_{ij_1}^{a_{11}} y_{ij_1}^{a_{12}} r_{ij_1}^{a_{13}} \right) \left(x_{ij_2}^{a_{21}} y_{ij_2}^{a_{22}} r_{ij_2}^{a_{23}} \right) \cdot \ldots \left(x_{ij_k}^{a_{k1}} y_{ij_k}^{a_{k2}} r_{ij_k}^{a_{k3}} \right) \quad \ldots(\text{A. } 2)$$

by the symmetric mean

$$< (a_{11}\, a_{12}\, a_{13}), (a_{21}\, a_{22}\, a_{23}), \ldots (a_{k1}\, a_{k2}\, a_{k3}) >_i{}'. \qquad (\text{A.}3)$$

The corresponding sample symmetric mean will be denoted similarly as in (A. 3) minus the prime ('). We shall find it expedient to operate with symmetric means in one portion of the ensuing derivation. Also let

$$\theta_{1_i} = \frac{N_i - n_i}{N_i - 1}$$

$$\theta_{2_i} = \frac{(N_i - n_i)(N_i - 2n_i)}{(N_i - 1)(N_i - 2)}$$

$$\theta_{3_i} = \frac{(N_i - n_i)\left(N_i^2 + N_i - 6N_i\, n_i + 6n_i^2 \right)}{(N_i - 1)(N_i - 2)(N_i - 3)}$$

and

$$\theta_{4i} = \frac{N_i (N_i - n_i)(N_i - n_i - 1)}{(N_i - 1)(N_i - 2)(N_i - 3)} \quad \ldots(\text{A. } 4)$$

From (A. 2) the variance can be expressed as

$$\begin{aligned} \text{Var}(y_c') = \sum_i w_i^2 \Big\{ & A_i^2\, \theta_{1i}\, \mu_{020_i}/n_i + B_i^2\, \text{Var}(\bar{r}_{n_i} \bar{x}_{n_i}) \\ & + \bar{x}_N^2\, \theta_{1i}\, \mu_{002i}/n_i + 2\bar{x}_N A_i\, \theta_{1i}\, \mu_{011i}/n_i - 2A_i B_i\, \text{Cov}(\bar{y}_{n_i}, \bar{r}_{n_i} \bar{x}_{n_i}) \\ & - 2\bar{x}_N B_i\, \text{Cov}(\bar{r}_{n_i} \bar{x}_{n_i}, \bar{r}_{n_i}) \Big\} \\ & + \text{Var}\Big(\sum_{i \neq j} w_i w_j\, \bar{r}_{n_i} \bar{x}_{n_j} \Big) \\ & - 2\text{Cov}\Big(\sum_i w_i A_i\, \bar{y}_{n_i}, \sum_{i \neq j} w_i w_j\, \bar{r}_{n_i} \bar{x}_{n_j} \Big) \\ & + 2\, \text{Cov}\Big(\sum_i w_i B_i\, \bar{r}_{ni} \bar{x}_{ni}, \sum_{i \neq j} w_i w_j\, \bar{r}_{ni} \bar{x}_{nj} \Big) \\ & - 2\bar{x}_N\, \text{Cov}\Big(\sum_i w_i\, \bar{r}_{ni}, \sum_{i \neq j} w_i w_j\, \bar{r}_{ni} \bar{x}_{nj} \Big). \end{aligned} \quad \ldots(\text{A. } 5)$$

We now find expressions for the terme in (A. 5). Details of the derivations are found in [1].

$$(a)\quad B_i^2 \operatorname{Var}(\bar{r}_{n_i}\,\bar{x}_{n_i}) = B_i^2 \Bigg[\frac{\theta_{1i}}{n_i}\Big(\mu_{200_i}\,\bar{r}^2_{N_i} + 2\mu_{101_i}\,\bar{x}_{N_i}\,\bar{r}_{N_i} + \mu_{002_i}\,\bar{x}^2_{N_i}\Big)$$

$$+\frac{2\theta_{2i}}{n_i^2}(\mu_{201_i}\,\bar{r}_{N_i}) + \frac{\theta_{4i}}{n_i^3}\mu_{202_i}$$

$$+\frac{\theta_{3i}(n_i-1)}{n_i^3}\mu_{200i}\,\mu_{002_i}$$

$$+\left\{\frac{2\theta_{3i}(n_i-1)}{n_i^3} - \frac{\theta^2_{1i}}{n_i^2}\right\}\mu^2_{101_i}\Bigg]$$

$$\rightarrow \frac{w_i^2}{(n_i-1)^2}\Bigg[n_i\Big(\mu_{200_i}\,\bar{r}^2_{N_i} + 2\mu_{101_i}\,\bar{x}_{N_i}\,\bar{r}_{N_i} + \mu_{002_i}\,\bar{x}^2_{N_i}\Big)$$

$$+ 2\mu_{201_i}\,\bar{r}_{N_i} + 2\mu_{102_i}\,\bar{x}_{N_i}$$

$$+\frac{n_i-1}{n_i}\mu_{200_i}\,\mu_{002_i}$$

$$+\frac{n_i-2}{n_i}\mu^2_{101_i} + \frac{1}{n_i}\mu_{202_i}\Bigg].$$

(b) $\bar{x}_N B_i \operatorname{Cov}(\bar{r}_{n_i}\,\bar{x}_{n_i},\ \bar{r}_{n_i})$

$$= \bar{x}_N B_i \left\{ \frac{\theta_{1_i}}{n_i}\Big(\bar{x}_{N_i}\,\mu_{002_i} + \bar{r}_{N_i}\,\mu_{101_i}\Big) + \frac{\theta_{2_i}}{n_i^2}\mu_{102_i} \right\}.$$

$$\rightarrow \frac{\bar{x}_N w_i}{n_i-1}\left\{ \bar{x}_{N_i}\mu_{002_i} + \bar{r}_{N_i}\mu_{101_i} + \frac{1}{n_i}\mu_{102_i} \right\}$$

$$(c)\ \operatorname{Var}\left(\sum_{i\neq j} w_i w_j \bar{r}_{n_i}\,\bar{x}_{n_j}\right)$$

$$= \sum_{i\neq j} w_i^2\ w_j^2 \Big\{\operatorname{Var}(\bar{r}_{n_i})\operatorname{Var}(\bar{x}_{n_j}) + \operatorname{Var}(\bar{r}_{n_i})\bar{x}^2_{N_i} + r^2_{N_i}\operatorname{Var}(\bar{x}_{n_j})$$

$$+ \operatorname{Cov}(\bar{x}_{n_i}, \bar{r}_{n_i})\operatorname{Cov}(\bar{x}_{n_i}, \bar{r}_{n_j}) + 2\operatorname{Cov}(\bar{x}_{n_i}, \bar{r}_{n_i})\bar{r}_{N_j}\bar{x}_{N_j}\Big\}$$

$$+ \sum_{i\neq j\neq k} w_i^2\ w_j w_k \{\operatorname{Var}(\bar{r}_{n_i})\bar{x}_{N_j}\bar{x}_{N_k}$$

$$+2\operatorname{Cov}(\bar{x}_{n_i}, \bar{r}_{n_i})\,\bar{r}_{N_j}\,\bar{x}_{N_k} + \operatorname{Var}(\bar{x}_{n_i})\,\bar{r}_{N_j}\,\bar{r}_{N_k}\}$$

$$-\left(\sum \frac{w_i^2\ \sigma^2_{ir}}{n_i}\right)\left(\sum \frac{w_i^2\ \sigma^2_{ix}}{n_i}\right) + \left(\sum \frac{w_i^2\ \sigma_{irx}}{n_i}\right)^2$$

$$+\sum \frac{w_i^3}{n_i}\left(\bar{x}^2_N\ \sigma^2_{ir} + 2\bar{x}_N\ \bar{r}_N\ \sigma_{irx} + \bar{r}_N\ \sigma^2_{ix}\right)$$

$$-2\sum \frac{w_i^3}{n_i}\left(\bar{x}_N\ \bar{x}_{N_i}\sigma^2_{ir} + \bar{r}_N\ \bar{x}_{N_i}\ \sigma_{irx} + \bar{r}_N\ \bar{r}_{N_i}\sigma^2_{ix} + \bar{x}_N\ \bar{r}_{N_i}\ \sigma_{irx}\right)$$

$$-\sum \frac{w_i^4}{n_i}\Big\{\frac{1}{n_i}\Big(\ \sigma^2_{ir}\ \ \sigma^2_{ix} + \sigma^2_{irk}\ \Big)$$

$$-x^2_{N_i}\ \sigma^2_{ir} - 2\bar{x}_{N_i}\ \bar{r}_{N_i}\ \sigma_{irx} - \bar{r}^2_{N_i}\ \sigma^2_{ix}\Big\}$$

$$(d)\ \operatorname{Cov}\left(\sum_i w_i A_i\ \bar{y}_{n_i},\ \sum_{i\neq j} w_i\ w_j\ \bar{r}_{n_i}\,\bar{x}_{n_j}\right)$$

$$= E\Big\{\sum_{i\neq j} w_i^2\ A_i\ w_j\ (y_{n_i}\ \bar{r}_{n_i}\ \bar{x}_{n_j} + \bar{y}_{n_i}\ \bar{x}_{n_i}\ \bar{r}_{n_j})$$

$$+ \sum_{i\neq j\neq k} w_i\ A_i\ w_j\ w_k\ \bar{y}_{n_i}\ \bar{r}_{n_j}\ \bar{x}_{n_k}\Big\}$$

$$-\left\{\sum_{i\neq j} w_i^2 A_i w_j (\bar{y}_{N_i}\bar{r}_{N_i}\bar{x}_{N_j}+\bar{y}_{N_i}\bar{x}_{N_i}\bar{r}_{N_j}) \right.$$

$$\left. \sum_{i\neq j\neq k} w_i A_i w_j w_k \bar{y}_{N_i}\bar{r}_{N_j}\bar{x}_{N_k}\right\}$$

$$=\sum_i w_i^2 A_2 \{\bar{x}_N \text{ Cov } (\bar{y}_{n_i}, \bar{r}_{n_i})+\bar{r}_N \text{ Cov } (\bar{x}_{n_i}, \bar{y}_{n_i})\}$$

$$-\sum_i w_i^3 A_i \{\bar{x}_{N_i} \text{ Cov } (\bar{y}_{n_i}, \bar{r}_{n_i})+\bar{r}_{N_i} \text{ Cov } (\bar{x}_{n_i}, \bar{y}_{n_i}\}$$

$$\to\sum \frac{w_i^2}{n_1}\left(1+\frac{w_i}{n_i-1}\right)(\bar{x}_N\,\sigma_{iyr})+\bar{r}_N\,\sigma_{ixy})$$

$$-\sum_i \frac{w_i^3}{n_i}\left(\frac{w_i}{n_i-1}\right)(\bar{x}_{N_i}\,\sigma_{iyr}+\bar{r}_{N_i}\,\sigma_{ixy}).$$

(e) $$\text{Cov}\left(\sum_i w_i B_i \bar{r}_{n_i}\bar{x}_{n_i}, \sum_{i\neq j} w_i w_j \bar{r}_{n_i}\bar{x}_{n_j}\right)$$

$$=\sum_{i\neq j} w_i^2 B_i w_j \left\{ E\left(\bar{r}_{n_i}^2\,\bar{x}_{n_i}\right)\bar{x}_{N_j}+E\left(\bar{r}_{n_i}\,\bar{x}_{n_i}^2\right)\bar{r}_{N_j}\right\}$$

$$+\sum_{i\neq j\neq k} w_i B_i w_j w_k E\,(\bar{r}_{n_i}\bar{x}_{n_i})\,\bar{r}_{N_j}\bar{x}_{N_k}$$

$$-\left[\sum_{i\neq j} w_i^2 B_i w_j \{E\,(\bar{r}_{n_i}\bar{x}_{n_i})\,\bar{r}_{N_i}\bar{x}_{N_j}+E\,(\bar{r}_{n_i}\bar{x}_{n_i})\,\bar{x}_{N_i}\bar{r}_{N_j}\}\right.$$

$$\left.+\sum_{i\neq j\neq k} w_i B_i w_j w_k E\,(\bar{r}_{n_i}\bar{x}_{n_i})\,\bar{r}_{N_j}\bar{x}_{N_k}\right]$$

$$=\sum_{i\neq j} w_i^2 B_i w_j \left\{\frac{\theta_{2i}}{n_i^2}\left(\mu_{102_i}\bar{x}_{N_j}+\mu_{201_i}\bar{r}_{N_j}\right)\right.$$

$$+\frac{\theta_{1i}}{n_i}\Big(\mu_{101_i}\bar{r}_{N_i}\bar{x}_{N_j}+\mu_{002_i}\bar{x}_{N_i}\bar{r}_{N_j}+\mu_{101_i}\bar{x}_{N_i}\bar{r}_{N_j}$$

$$\left.+\mu_{200_i}\bar{r}_{N_i}\bar{r}_{N_j}\Big)\right\}$$

$$\rightarrow \bar{x}_N \sum_i \frac{w_i^3}{n_i-1}\left(\frac{\mu_{102_i}}{n_i}+\mu_{101_i}\bar{r}_{N_i}+\mu_{002_i}\bar{x}_{N_i}\right)$$

$$+\bar{r}_N \sum \frac{w_i^3}{n_i-1}\left(\frac{\mu_{201_i}}{n_i}+\mu_{101_i}\bar{x}_{N_i}+\mu_{200_i}\bar{r}_{N_i}\right)$$

$$-\sum_i \frac{w_i^4}{n_i-1}\left\{\frac{1}{n_i}\left(\mu_{102_i}\bar{x}_{N_i}+\mu_{201_i}\bar{r}_{N_i}\right)\right.$$

$$\left.+2\mu_{101_i}\bar{r}_{N_i}\bar{x}_{N_i}+\mu_{002_i}\bar{x}^2_{N_i}+\mu_{200_i}\bar{r}^2_{N_i}\right\}.$$

(*f*) $\bar{x}_N \,\mathrm{Cvo}\left(\sum_i w_i \bar{r}_{ni}{}', \sum_{i\neq j} w_i w_j \bar{r}_{ni}\bar{x}_{nj}\right)$

$$=\bar{x}_N \sum_{i\neq j} w_i^2 w_j \{\mathrm{Var}(\bar{r}_{ni})\bar{x}_{N_j}+\mathrm{Cov}(\bar{r}_{ni}, \bar{x}_{ni}\bar{r}_{Nj})$$

$$-\bar{x}_N \sum_i \frac{w_i^2}{n_i}(\bar{x}_N \mu_{002i}+\bar{r}_N \mu_{101i})$$

$$-\bar{x}_N \sum_i \frac{w_i^3}{n_i}(\bar{x}_{Ni}\mu_{002i}+\bar{r}_{Ni}\mu_{101i})$$

(*g*) $A_i B_i \,\mathrm{Cov}(\bar{y}_{ni}, \bar{r}_{ni}\bar{x}_{ni})$

$$=A_iB_i\,[E\{<(100)>_i <(010)>_i <(001)>_i\}$$
$$-<(010)>_i' E\{<(100)>_i <(001)>_i\}]$$

$$=A_i B_i\left[\frac{1}{n_i^2}\{<(020)>_i'+(n_i-1)(<(100); (011)>_i'\right.$$
$$+<(110), (001)>_i'+<(010), (010)>_i'$$
$$+(n_i-1)(n_i-2)<(100), (010), (001)>_i'\}$$
$$-\frac{1}{n_i N_i}\{<(020)>_i'+(N_i-1)<(010), 010)>_i'$$
$$+(n_i-1)(<(110), (001)>_i'+<(100); (011)>_i'$$
$$\left.+(n_i-1)(N_i-2)<(100), (010), (001)>_i'\}\right]$$

$$=A_i B_i\left(\frac{N_i-n_i}{N_i n_i}\right)\left[\frac{1}{n_i}\{<(020)>_i'-<(010), (010)>_i'\}\right.$$
$$+\left\{1-\frac{1}{n_i}\right\}\{<(110), (001)>_i'+<(100), (011)>_i'$$
$$\left.-2<(100), (010), (001)>_i'\}\right].$$

Using (3.3), we find after some simplification,

$$<(020)>_i' - <(010),(010)>_i' = \frac{N_i}{N_i-1}\,\mu_{020_i}$$

and

$$<(110),(001)>_i' + <(100),(011)>_i' - 2<(100),(010),(001)>_i'$$
$$= \frac{1}{(N_i-1)(N_i-2)}(N_i^2\,\bar{x}_{N_i}\,\mu_{011_i} + N_i^2\,\bar{r}_{N_i}\,\mu_{110_i}$$
$$-2N_i\,\mu_{020_i}).$$

Therefore,

$$A_i\,B_i\,\mathrm{Cov}\,(\bar{y}_{n_i},\,\bar{r}_{n_i}\,\bar{x}_{n_i})$$
$$= A_i\,B_i\,\frac{\theta_{1_i}}{n_i^2}\Big\{\,\theta_{2_i}\,\mu_{020_i}\,\frac{N_i(n_i-1)}{N_i-2}\,(\bar{x}_{Ni}\,\mu_{011_i}$$
$$+\bar{r}_{Ni}\,\mu_{110_i})\Big\}$$
$$\to\left(1+\frac{w_i}{n_i-1}\right)\left(\frac{w_i}{n_i-1}\right)\left\{\frac{\mu_{020_i}}{n_i}+\left(\frac{n_i-1}{n_i}\right)\right.$$
$$\left.(\bar{x}_{Ni}\,\mu_{011_i}+\bar{r}_{Ni}\,\mu_{110_i})\right\}.$$

We substitute the results (*a*)—(*g*) in (A. 5). After some simplifying and grouping of terms we obtain

$$\mathrm{Var}\,(y'_c)\to\sum\frac{w_i^2}{n_i}\left(\bar{r}_N^2\,\mu_{120_i}-2\bar{r}_N\,\mu_{110_i}+\mu_{200_i}\right)$$

$$+2\sum\frac{w_i^3}{n_i(n_i-1)}\left(\bar{r}_N\,\mu_{201_i}+\bar{r}_N\,\bar{r}_{N_i}\,\mu_{200_i}+\bar{r}_N\,\bar{x}_{N_i}\,\mu_{101_i}-\bar{r}_N\,\mu_{110_i}\right)$$

$$+\sum\frac{w_i^4}{n_i(n_i-1)}\left(2\bar{x}_{N_i}\,\mu_{011_i}+2\bar{r}_{N_i}\,\mu_{110_i}+2\mu^2_{101_i}\right.$$
$$\left.+\mu_{200_i}\,\mu_{002_i}-2\bar{x}_{N_i}\,\mu_{102_i}-2\bar{r}_{N_i}\,\mu_{201_i}\right)$$

$$+\sum\frac{w_i^4}{(n_i-1)^2}\left(2\bar{r}_{N_i}\,\mu_{201_i}+2\bar{x}_{N_i}\,\mu_{102_i}-\mu_{101_i}^2\right.$$
$$\left.-2\bar{x}_{N_i}\,\mu_{011_i}-2\bar{r}_{N_i}\,\mu_{110_i}\right)$$

$$+\sum\frac{w_1^4}{n_i(n_i-1)^2}\left(\mu_{202_i}-\mu_{020_i}+2\bar{x}_{N_i}\,\mu_{011_i}\right.$$
$$\left.+2\bar{r}_{N_i}\,\mu_{110_i}+\bar{x}_{N_i}^2\,\mu_{002_i}+\bar{r}_{N_i}^2\,\mu_{200_i}+2\bar{x}_{N_i}\,\bar{r}_{Ni}\,\mu_{101_i}\right)$$

$$+\left(\sum \frac{w_i^2\ \mu_{200_i}}{n_i}\right)\left(\sum \frac{w_i^2\ \mu_{002_i}}{n_i}\right)$$

$$-\sum \frac{w_i^4}{n_i^2}\mu_{200_i}\ \mu_{002_i} + \left(\sum \frac{w_i^2\ \mu_{101_i}}{n_i}\right)^2$$

$$-\sum \frac{w_i^4\ \mu_{101_i}}{n_i^2}.$$

To simplify (A. 6) furtner we express the higher moments $\mu_{201_i}\ \mu_{102_i}$ and μ_{202_i} in terms of the moments of lower order. With some work it can be shown that

$$\mu_{201_i} = \mu_{110_i} - \bar{r}_{N_i}\ \mu_{200_i} - \bar{x}_{N_i}\left(\ \bar{y}_{N_i} - \bar{x}_{N_i}\ \bar{r}_{N_i}\right)$$

$$\rightarrow \mu_{110_i} - \bar{r}_{N_i}\ \mu_{200_i} - \bar{x}_{N_i}\ \mu_{101_i}$$

$$\mu_{102_i} \rightarrow \mu_{011_i} - \bar{x}_{N_i}\ \mu_{002_i} - \bar{r}_{N_i}\ \mu_{101_i}.$$

and

$$\mu_{202_i} \rightarrow \mu_{020_i} + \mu_{101_i}^2 + 2\bar{x}_{N_i}\ \bar{r}_{N_i}\ \mu_{101_i}$$

$$-2\bar{r}_{N_i}\ \mu_{110_i} + \bar{r}_{N_i}^2\ \mu_{200_i} - 2\bar{x}_{N_i}\ \mu_{011_i} + \bar{x}_{N_i}^2\ \mu_{002_i}.$$

Finally upon substituting these three relations in (A. 6) and after further simplification, we get the variance form (3.2).

Summary

A combined unbiased ratio-type estimator which is an analogue of the Hartley-Ross estimator in stratified sampling is presented and its variance is derived. Using actual data, it is found that the former estimator usually is more efficient in real situations where ratio-type estimators are applicable. Also, it is seen that the use of the large-sample variances of these estimators can lead to gross underestimation even for moderately large samples.

REFERENCES

[1] David, I. O.,(1917). : *Contributions to Ratio Method of Estimations*, Ph.D Thesis, Ames, Iowa, Library, Iowa State University.

[2] Goodman, L.A. and Hartley, H.O. (1958). : The Precision of Unbiashed Ratio-Type Estimators, *Journal of the American Statistical Association*, 53 : 491-508.

[3] Hartley, H.O. and Ross, A. (1954). : Unbaised Ratio Estimators *nature*, 174 : 270-271

[4] Lahiri, D.B. (1959). : A Method of Sample Selection Provid ng Unbaised Ratio Estimates, *Bulleetin Institute International de Statistique*, 33 : 133-140.

[5] Nanjamma, N.S., Murthy, M.N. and Sethi, N.S. (1959). : Some Sampling Systems Providing Unbaised Ratio Estimators, *Sankhya*, 21 : 199-214.

[6] Nieto De Pascual, J. (1961) : Unbaised Ratio Estimators in Stratified Sampling, *Journal of the American Statistical Association*, 56 : 70-87. 1961.

[7] Rao, J.N.K. (1964). : Unbased Ratio and Regression Estimators in Multi-state Sampling, *Journal of the Indian Society of Agricultural Statistics*, 16 : 175-188.

[8] Robson, S.D. (1957). : Applications of Multivariate Polykays to the Theory of Unbaised Ratio-type Estimation *Journal of the American Statistical Association*, 52 : 511-522.

[9] Ross, A. (1960). : *On Two Problems in Sampling Theory : Unbiased Ratio Estimators and Variance Estimates in Optimum Sampling Designs*, PH.D. Thesis, Ames, Iowa, Library, Iowa State University.

[10] Tukey, J.W. : Keeping Moment-Like Sampling Computations Simple, *Annals of Mathematical Statistics*, 27 : 37-54.