MULTIVARIATE UNBIASED RATIO-TYPE ESTIMATION IN FINITE POPULATION

BY

V. RAMACHANDRAN* AND S.S. PILLAI**

(Received in October, 1969;
Revised in August, 1974)

1. INTRODUCTION

Estimators using information on auxiliary variates have been developed by several authors for achieving greater precision in estimating population means of a particular main variate as compared to the estimators based on the sample mean. The biased ratio estimator is one among them. Hartley and Ross (1954) gave an unbiased ratio estimator by using information on one auxiliary variate. Olkin (1958) extended the biased ratio method of estimation to the case when information on several auxiliary variates is available.

This paper gives an unbiased ratio estimator of population mean using data on two or more auxiliary variables along with its variance. It also indicates how to obtain the estimator of the variance. The utility of the results obtained is illustrated with the help of the data collected from a sample survey on pepper crop.

2. MULTIVARIATE UNBIASED RATIO-TYPE ESTIMATION

Let the information on p auxiliary variates x_1, x_2—x_p for each of the N units of the population be available and the observation on the main variate be Y. Let μ_{xi} denote the known population mean of the ith auxiliary variate for $i=1, 2...p$. Let a simple random sample of size n be drawn without replacement and the outcome for the main and auxiliary variates be Y_j, X_{1j} $X_{2j}...X_{pj}$, R_{1j}, R_{2j}— $j=1, 2...n$, where $R_{ij}=Y_j/X_{ij}$, $i=1,2, 3...p$.

---

*Indian Institute of Management, Ahmedabad.

**Institute of Agricultural Research Statistics. New Delhi.

Before giving the main result in Theorem 2.1 below, we first present the following lemma:

Let
$$\bar{R}_i = \frac{1}{n}\sum_{j=1}^{n} Y_j/X_{ij}\,;$$
$$\bar{Y} = \frac{1}{n}\sum_{j=1}^{n} Y_j$$
$$\bar{X}_i = \frac{1}{n}\sum_{j=1}^{n} X_{ij}$$
and
$$\bar{Y}_i' = \bar{R}_i \mu_{xi} + \frac{n\,(N-1)}{N\,(n-1)}\,(\bar{Y}_i - \bar{R}_i \bar{X}_i)$$

where $\bar{Y}_i'$ is the Hartley and Ross's unbiased ratio estimator for the population mean of y when auxiliary information on variate x_i is used and μ_{xi} is the population mean for the variate x_i.

Lemma 2.1

The covariance between $\bar{Y}'_i$ and $\bar{Y}_{i'}'$ $i \neq i'$ for $i, i' = 1, 2 \ldots p$ is given by

$$\text{Cov}\left(\bar{Y}_i'\;\bar{Y}'_{i'}\right) = \frac{1}{n}\left\{\bar{r}_i\,\bar{r}_{i'}\;\sigma x_i\,x_i' + \sigma yy - \bar{r}_i\;\sigma x_i y - \bar{r}_{i'}\;\sigma x_i' y\right\}$$
$$+\frac{1}{n\,(n-1)}\;\{\sigma x_i x_i'\;\;\sigma R_i R_i' + \sigma x_i R_i'\;\sigma x_i' R_i\} \qquad (2.1)$$

Where

$$\bar{r}_i \text{ stands for } \frac{1}{N}\sum_{i=1}^{N} R_{ij} \text{ and } N \text{ is very large.}$$

For $i' = i$ this expression gives the variance of $\bar{Y}'_i$ (Robson, 1957, Page 519).

Proof:

For simplicity of presentation, we give the proof for $p=2$ only.

In terms of symmetric mean notation (Tukey, 1956) we have

$\bar{R}_1 = \langle(00010)\rangle$

$\bar{R}_2 = \langle(00001)\rangle$

$$\bar{X}_1 = <(01000)>$$

$$\bar{X}_2 = <(00100)>$$

$$\bar{Y} = <(10000)>$$

$$\bar{Y}_1 = <(00010)><(01000)>' - \frac{N-1}{N}\{<(01000)><(00010)> - <(10000)>\}$$

$$\bar{Y}_2 = <(00001)><(00100)>' - \frac{N-1}{N}\{<(00100)><(00001)> - <10000)>\}$$

Where the symmetric means are defined for the five variate population $(Y_i, X_{1i}, X_{2i}, R_{1i}, R_{2i})$; $i=1, 2, 3, \ldots N$.

An angular bracket $<>'$ With a prime denotes the symmmetric mean for the population, while $<>$ without the prime denotes sample symmetric mean.

$$\bar{Y}_1'\bar{Y}_2' = \{<(01000)>'<(00010)> - \frac{N-1}{N}[<(01000)><(00010)> - <(10000)>]\} \times \{<(00001)><(00100)>' - \frac{N-1}{N}[<(00100)><(00001)> - <(10000)>]\}.$$

Noting that Cov $(\bar{Y}_1' \bar{Y}_2') = E(\bar{Y}_1 \bar{Y}_2') - \mu_y^2$ where μ_y is the population mean for the Y-variate. By using Robsons (1957, Page 513) method of multiplication of symmetric means, we have

$$\begin{aligned}\text{Cov}\,(\bar{Y}_1' \bar{Y}_2') = \frac{(N-1)^2}{N^2 n\,(n-1)} \{&<(01100)><(00011)>' + <(01001)<(00110)> \\ &+ (n-2)<(01100), (00010), (00001)>' + (n-2)<(01001), (00100), (00010)>' \\ &+ (n-2) <(01000), (00110), (00001)>' + (n-2) <(01001), (00010), (00100)>' \\ &+ (n-2)(n-3)<(01000), (00100), (00010), (00001)>'\} \\ + \frac{(N-1)^2}{N^2 n}\{&<(10000)>' + (n-1)<(10000), (01000>' - <(11000), (00010)>' \\ &- <(01000), (10010)>' - (n-2)<(10000), (01000), (00010)>' \\ &- <(10100), (00001)>' - <(00100), (10001)>' - (n-2) <(10000), (00100), (00001>'\end{aligned}$$

$$+\frac{1}{N^2n}\{<(01111)>'+(N-1)<(01100),(00011)>'+(n-1)<(01110),(00001)>'$$

$$+(n-1)<(01101),(00010)>'+(n-1)(n-2)<(01100),(00010),(00001)>'$$

$$+(N-1)<(00111),(01000)>'+(N-1)<(01011),(00100)>'$$

$$+(N-1)(N-2)<(00011),(01000),(00100)>'+(n-1)<(01010),(00101)>'$$

$$+(n-1)<(01001),(00110)>'+(n-1)(N-2)<(01010),(00100),(00001)>'$$

$$+(n-1)(N-2)<(01001),(00100),(00010)>'+(n-1)(N-2)<(01000),(00100),(00101)>'$$

$$+(n-1)(N-2)<(01000),(00110),(00001)>'+(n-1)(N-2)(N-3)<(01000),(00100),(00001)>'\}$$

$$+\frac{N-1}{N^2n}\{<(11010)>'+\frac{1}{N}<(00101)>'+(N-1)<(10010),(01000)>'+(N-1)<(10001),(00100)>'\}$$

$$+\frac{(N-1)(n-1)}{N^2n}\{<(11000),(00010)>'+<(01010),(10000)>'+<(10100),(00001)>'$$

$$+<(00101),(10000)>'+(N-2)<(10000),(01000),(00010)>'+(N-2)<(10000),(00100),(00001)>'\}$$

$$-\frac{N-1}{N^2n}\{<(01110),(00001)>'+<(00110),(01001)>'+<(01011),(00100)>'$$

$$+2<(01100),(00011)>'+<(01101),(00010)>'+<(01001),(00110)>'$$

$$+<(00111),(01000)>'+(N-2)<(00110),(01000),(00001)>'$$

$$+2(N-2)<(01001),(01000),(00100)>'+(N-2)<(01001),(00100),(00010)>'\}$$

$$-\frac{(N-1)(n-2)}{N^2n}\{<(01010),(00100),(00001)>'+2<(01100),(00010),(00001)>'$$

$$+<(01001),(00100),(00010)>'+<(00110),(01000),(00001)>'$$

$$+<(00101),(01000),(00010)>'$$

$$+2(N-3)<(01000),(00100),(00010),(00001)>'\}$$

$$-\frac{1}{N}\{<(10000)>'-(N-1)<(10000),(10000)>'\}$$

An unbiased estimator of Cov $(\bar{Y}_1' \bar{Y}_2')$ can be obtained from (2.2) by replacing the population symmetric means with respective sample symmetric means. We assume that N is very large and consequently omit terms of the order of $1/N$. Further for N large, we can use the approximation :

$$<(a_1)(a_2)\ldots(a_n)>' = <(a_1)>'<(a_2)>'\ldots<(a_n)>'$$

Then, the expresstion for Cov $(\bar{Y}_1' \bar{Y}_2')$ reduces to

$$\text{Cov}(\bar{Y}_1'\bar{Y}_2') = \frac{1}{n}\{\bar{\gamma}_1 \bar{\gamma}_2 \, \sigma x_1x_2 + \sigma yy - \bar{\gamma}_1 \sigma x_1 y - \bar{\gamma}_2 \, \sigma x_2 y\}$$
$$+ \frac{1}{n\,(n-1)}\{\sigma x_1x_2 \, \sigma R_1R_2 + \sigma x_1R_2 \, \sigma x_2R_1\} \qquad \ldots(2.3)$$

Evidently for general p, (2.3) will take the form as in (2.1).

The unbiased estimator of Cov $(\bar{Y}_i' \; \bar{Y}'_i)$ can be obtained on the same lines followed by Goodman and Hartley (1958) [vide (35), Page 499] in obtaining unbiased estimator for Var (Y'_i). $\bar{Y}_i'$

We can define the multivariate unbiased ratio-type estimator of the population mean as

$$\hat{\bar{Y}}_r \; (MV) = \sum_{j=1}^{p} w_j \; \bar{Y}_j' \text{ Where}$$

$\bar{Y}_j'$ is the Hartley and Ross's unbiased ratio estimator and weights w_i are chosen such that their sum is 1. We state the following theorem :

Theorem 2.1. The optimum value of w_i is given by $\frac{R_i}{D}$ where R_i is the sum of all the elements in ith row of the matrix V^{-1} and D is the sum of all the elements in V^{-1}, V being the varinance-covariance matrix with elements v_{ij} given by (2.1.). The minimum variance of $\hat{\bar{Y}}_r$ (MV) is given by $1/D$. The proof of the theorem is on the lines given in article 4 of Olkin (1958).

3. Numerical Illustration

The utility of the results obtained in the preceding section will now be illustrated with the help of data collected from a sample survey conducted on pepper in the State of Kerala during the years 1966-68.

The data relates to the number of pepper standards and area under pepper enumerated completely in each of a simple random sample of 120 villages drawn independently during 1966-67 and 1967-68, without replacement from the population of 3280 villages.

Information on two auxiliary variables: (1) Geographical area of the village, and (2) Dry land area of the village was available for all villages in the State.

The estimates of the average number of pepper standards and the average area under pepper per village for the State were obtained by using the information on above variates in them.

These estimates were also obtained for each year using other procedures. The different estimates are given in the tables 1-4 in the appendix. The results have been given for both the years and for two characters to highlight the consistency in the results obtained.

In the notations indicating the type of estimator used (Column 2 of different tables) the first suffix stands for the character for which the estimate is made; *viz.*, "1" for number of pepper standards and "2" for area under pepper. The second suffix stands for biased (R) or unbiased ratio type (r) estimator. The third suffix denotes the auxiliary variable used in building up the ratio estimator. For instance, $\hat{\bar{Y}}_1 R_1$ stands for the biased ratio estimator for number of pepper standards using geographical as area auxiliary variable $\hat{\bar{Y}}_{2r2}$ stands for the Hartley-Ross' unbiased ratio type estimator for area under pepper using dry-land area as the auxiliary variate. The notation $\hat{\bar{Y}}_{1R}$ (MV) denotes the bivariate biased ratio estimator (Olkin) for number of pepper standards, and $\hat{\bar{Y}}_{1r}$ (MV) represents the bivariate unbiased ratio-type estimator (2.8) of the number of pepper standards in the State. $\hat{\bar{Y}}_1$ (sm) represents the estimator based on the simple mean.

From the tables it may be seen that the multivariate unbiased ratio-type estimator, while being unbiased, is almost equally efficient as the multivariate biased ratio estimator and is in all cases more efficient than each of the single unbiased ratio estimators. It is also in all cases more efficient than the estimator based on the simple mean.

ACKNOWLEDGEMENTS

The authors are extremely grateful to Dr. Daroga Singh, Director, Institute of Agricultural Research Statistics, for his keen interest and encouragement in the preparation of the paper. They also thank the referee for his comments which helped in improving its presentation.

APPENDIX

TABLE 1

Estimates of average number of pepper standards per village in Kerala State during 1966-67 obtained by various estimation procedures, along with their variances and relative biases

Sl. No.	*Estimator*	*Estimate of average no. of standards per village*	*Variance*	*% gain in precision over simple mean*	*% bias relative to estimate*
(1)	(2)	(3)	(4)	(5)	(6)
1.	$\hat{\bar{Y}}_1$ (sm)	25,200	1,95,14,801	—	—
2.	$\hat{\bar{Y}}_{1R1}$	28,737	1,06,38,999	83.4	0.7301
3.	$\hat{\bar{Y}}_{1R2}$	21,787	1,10,60,294	76.4	1.2938
4.	$\hat{\bar{Y}}_{1r1}$	29,242	1,19,73,724	62.9	—
5.	$\hat{\bar{Y}}_{1r2}$	21,002	1,48,73,373	31.2	—
6.	$\hat{\bar{Y}}_{1R}$ (MV)	28,096	95,34,271	104.6	0.7804
7.	$\hat{\bar{Y}}_{1r}$ (MV)	30,195	1,19,42,231	63.4	—

TABLE 2

Estimates of average number of pepper standards per village in Kerala State during 1967-68 obtained by various estimation procedures along with their variances and biases.

Sl. No.	*Estimator*	*Estimate of average no. of standards per village*	*Variance*	*% gain in precision over simple mean*	*% bias relative to estimate*
(1)	2)	(3)	(4)	(5)	(6)
1.	$\hat{\bar{Y}}_1$ (sm)	16,384	69,75,448	—	—
2.	$\hat{\bar{Y}}_{1R1}$	26,868	45,95,802	51.8	.4023
3.	$\hat{\bar{Y}}_{1R_2}$	23,194	21,61,201	227.7	—.4013
4.	$\hat{\bar{Y}}_{1r1}$	27,978	48,37,874	44.5	—
5.	$\hat{\bar{Y}}_{1r2}$	23,324	21,50,652	224.3	—
6.	$\hat{\bar{Y}}_{1R}$ (MV)	23,194	21,22,555	228.2	—.4122
7.	$\hat{\bar{Y}}_{1r}$ (MV)	22,586	20,99,536	232.2	—

TABLE 3

Estimates of average area under pepper per village in Kerala State during 1966-67 obtained by various estimation procedures along with their variances and relative biases

Sl. No.	*Estimator*	*Estimated average area (in acres) per village*	*Variance*	*% gain in precision over simple mean*	*% bias relative to mean*
(1)	(2)	(3)	(4)	(5)	(6)
1.	$\hat{\bar{Y}}_2$ (sm)	632.01	14,595	—	—
2.	$\hat{\bar{Y}}_{2r1}$	720.73	3,980	266.7	−.1767
3.	$\hat{\bar{Y}}_{2R2}$	546.44	3,507	316.1	.2538
4.	$\hat{\bar{Y}}_{2r1}$	719.94	4,013	263.8	—
5.	$\hat{\bar{Y}}_{2r2}$	543.31	3,598	305.6	—
6.	$\hat{\bar{Y}}_{2r}$ (MV)	524.11	3,502	316.7	.3089
7.	$\hat{\bar{Y}}_{2r}$ (MV)	574.72	3,578	307.9	—

TABLE 4

Estimates of average area under pepper per village in Kerala State during 1967-68 obtained by various estimation procedures along with their variances and relative biases

Sl. No.	*Estimator*	*Estimated average area (in acres) per village*	*Variance*	*% gain in precision over simple mean*	*% bias relative to mean*
(1)	(2)	(3)	(4)	(5)	(6)
1.	$\hat{\bar{Y}}_2$ (sm)	376.81	2,862	—	—
2.	$\hat{\bar{Y}}_{2R1}$	617.94	2,122	34.8	.6655
3.	$\hat{\bar{Y}}_{2R2}$	533.44	687	316.5	—.2127
4.	$\hat{\bar{Y}}_{2r1}$	628.01	2,187	30.8	—
5.	$\hat{\bar{Y}}_{2r2}$	525.54	716	299.7	—
6.	$\hat{\bar{Y}}_{2R}$ (M.V.)	515.55	647	342.3	—.3986
7.	$\hat{\bar{Y}}_{2r}$(M.V.)	499.18	652	338.9	—

REFERENCES

1. Hartley, H.O. and Ross, A. (1954) : 'Unbiased ratio estimators', Nature 174 (1954), 270-271.
2. Goodman, L.A. and Hartley, H.O. (1958) : 'The precision of unbiased ratio-type Estimators'. JASA 53 (1958), 491-508.
3. Tukey, John W. (1956) : 'Keeping moment-like sampling computations simple'. Annals of Mathematical Statistics, 27, (1956), 37-54.
4. Robson, D.S. (1957) : 'Applications of Multivariate Polykays to theory of unbiased ratio-type estimation'. JASA Vol. 52 (1957), 511-522.
5. Olkin, I. (1958) : 'Multivariate ratio estimation for finite population'. Biometrika, 43 (1958), 154-63.