ON SAMPLING WITH VARYING PROBABILITIES

BY

A. K. SRIVASTAVA

Dissertation submitted in fulfilment of the requirements for the award of Diploma in Agricultural and Animal Husbandry Statistics of the Institute of Agricultural Research Statistics

(I.C.A.R.)

NEW DELHI

1967

ACKNOWLEDGEMENTS

I have pleasure in expressing my deepest sense of gratitude to Dr. B.V. Sukhatme, Senior Professor of Statistics, Institute of Agricultural Research Statistics(I.C.A.R.) for his valuable guidance, constant encouragement and constructive criticism during the course of investigation and presentation of this thesis.

I am grateful to Dr. G.R. Seth, Statistical Adviser, Institute of Agricultural Research Statistics(I.C.A.R.) for providing necessary facilities for research work.

I am also thankful to Shri M.S. Avadhani, Statistician, Indian Society of Agricultural Statistics and Shri Ravindra Singh, Research Scholar, Institute of Agricultural Research Statistics for helping me in presenting this thesis.

A.K. Srivastava

(A.K. Srivastava)

CONTENTS

INTRODUCTION: - Consider a finite population model of size N and let y_t(t = 1, 2, ... N) be the y-characteristic values of the population. The main problem of sampling theory is to estimate the total $Y = y_1 + y_2 + \ldots + y_N$ with the help of a sample of observations drawn from the population. When information x_t on an auxiliary character x is available for all t = 1, 2, ... N and x and y are highly correlated, it is customary to use these x_t's either in the selection of the sample or for estimation after selecting the sample by simple random sampling, to obtain a better estimate of Y. One way of using x_t's in selecting the sample is to draw the units from the population one after another and with probability proportional to x_t's (t=1, 2, ..., N) either with or without replacement. This procedure of selection is generally known as the probability proportional to size (p. p. s) sampling.

In the case of p. p. s sampling without replacement a number of estimation procedures were developed. In this connection reference may be made to the papers by Horvitz and Thompson (1952), Narain(1951), Yates and Grundy (1953), Des Raj(1956), Murthy(1957), Das(1958), Hartley and Rao(1962). However, these are complicated and less useful for application in practice, specially when the sample size is large. Consequently, Rao, Hartley and Cochran(1962) have proposed a simple technique of unequal probability sampling without replacement.

The main advantages of this methodology, besides the simplicity of selection, are:

(i) It provides an estimate of the population total which is always more efficient than the standard estimator in sampling with unequal probability and with replacement.

(ii) It does not entail heavy computations even for the sample size $n > 2$ for finding out the estimate of the population total and its variance.

(iii) It furnishes an unbiased estimate of the variance which is always positive,

(iv) It provides an exact formula for the variance of the estimate of Y.

In actual practice, it may happen that sometimes some observations may be missing due to unforeseen circumstances. For example, in agricultural yield surveys the enumerator may be accidentally held up and may not be able to contact a unit before the harvest time and consequently may omit it. In such situations the applicability of Rao, Hartley and Cochran technique is hampered since the estimate proposed by these authors does not remain unbiased, and even though an unbiased estimate can be obtained when some observations are missing its variance increases considerably. Theref-ore it is pertinent to examine the feasibility of modifying the probability sampling scheme under consideration so as to enhance its practical applicability .

An attempt has been made here to overcome this difficulty. Two modifications of the R.H.C. scheme have been discussed together with the estimation methodology. These are found to be better than the R.H.C. estimate in some situations. Also, the ordinary ratio estimate is compared with the R.H.C. estimate with and without missing observations.

Section I

ON RAO-HARTLEY - COCHRAN PROCEDURE OF UNEQUAL PROBABILITY SAMPLING.

1.1. Introduction

To facilitate further reference and ensure completeness the Rao-Hartley and Cochran procedure of unequal probability sampling wihout replacement(R.H.C. scheme) is briefly described in the following subsection(1.2). Two different modifications of the R.H.C. scheme have been introduced and their efficiencies have been compared with some estimates of unequal probability sampling in subsequent subsections(1.3, 1.4 and 1.5).

1.2. The Rao-Hartley and Cochran sampling procedure:

For selecting a sample of size n units from the population in question, the R.H.C. scheme consists of the following two steps:

(a) Split the population at random into n groups of sizes $N_1, N_2, \ldots, N_n$ such that $N_1+N_2+\ldots+N_n = N$.

(b) Draw a sample of size one with probability proportional to p_t from each of these n group independently.

where $p_t = x_t / X$ such that $X = \sum_{t=1}^{N} x_t$

It is shown that [cf. (1)] the statistic

$$\hat{Y} = \sum_{i=1}^{n} \frac{y_i}{p_i/\pi_i} \quad \ldots . \quad (1.2.1)$$

where π_i is the sum of the probabilities of the units falling in the i th group and the suffixes 1, 2, ..., n denote the n units selected from the n groups separately, is an unbiased estimate of Y and its variance is

$$V(\hat{Y}) = \frac{\sum_{i=1}^{n} N_i(N_i-1)}{N(N-1)} \sum_{t=1}^{N} p_t \left(\frac{y_t}{p_t} - Y \right)^2 \quad \ldots (1.2.2)$$

In case N is a multiple of n, and $N_1 = N_2 = \ldots = N_n$, then $V(\hat{Y})$ is minimum and its value is

$$\text{Min } V(\hat{Y}) = \frac{N-n}{n(N-1)} \sum_{t=1}^{N} p_t \left(\frac{y_t}{p_t} - Y \right)^2 \quad \ldots (1.2.3)$$

Again, when N is not a multiple of n, say $N = nR + k$ where $0 < k < n$ and R is a positive integer and

$N_1 = N_2 = \ldots = N_k = R + 1$; $N_{k+1} = N_{k+2} = \ldots N_n = R$.

then (1.2.2) reduces to

$$V(\hat{Y}) = \frac{1}{n} \left[1 - \frac{n-1}{N-1} + \frac{k(n-k)}{N(N-1)} \right] \sum_{t=1}^{N} p_t \left(\frac{y_t}{p_t} - Y \right)^2 \quad \ldots (1.2.4)$$

An unbiased estimate of $V(\hat{Y})$ given by (1.2.4) is

$$v(\hat{Y}) = \frac{N^2 + k(n-k) - Nn}{N^2(n-1) - k(n-k)} \sum_{i=1}^{n} \pi_i \left(\frac{y_i}{p_i} - Y \right)^2 \quad \ldots (1.2.5)$$

It may be remarked that this estimate is always positive. An estimate of Min $V(\hat{Y})$ is obtained from (1.2.5) by setting $k = 0$. Thus,

$$\text{Min } v(\hat{Y}) = \frac{1}{n-1} \left(1 - \frac{n}{N} \right) \sum_{i=1}^{n} \pi_i \left(\frac{y_i}{p_i} - Y \right)^2 \quad \ldots (1.2.6)$$

1.3. An alternative scheme:

Let n be even. The scheme consists of two steps:

(a) Split the population at random into $n/2$ groups of sizes $N_1, N_2, \ldots, N_{n/2}$ such that $\sum_{i=1}^{n/2} N_i = N$.

(b) Draw a sample of size 2 with probability proportional to p_t's and with replacement from each of these groups independently.

<u>Estimator of the population total Y:</u>

Define

$$\hat{Y}_1 = \sum_{i=1}^{n/2} \frac{1}{2} \left(\frac{y_{i1}}{p_{i1}/\pi_i} + \frac{y_{i2}}{p_{i2}/\pi_i} \right) \qquad \ldots \quad (1.3.1)$$

where y_{i1} and y_{i2} are the y-characteristic values of the two units selected in the i th group; p_{i1} and p_{i2} are the corresponding initial probabilities of the units and

$$\pi_i = \sum_{\text{Group } i} p_j \ .$$

$$\text{Now, } E(\hat{Y}_1) = E_1 \ E_2(\hat{Y}_1)$$

where E_2 is the expectation over a given split and E_1 is the expectation over all possible splits of the population into $n/2$ groups of sizes $N_1, N_2, \ldots, N_{n/2}$. But

$$E_2(\hat{Y}_1) = \sum_{i=1}^{n/2} E_2\left(\frac{y_{i1}}{p_{i1}/\pi_i}\right) = \sum_{i=1}^{n/2} Y_i = Y$$

$$\text{since } E_2\left(\frac{y_{i1}}{p_{i1}/\pi_i}\right) = E_2\left(\frac{y_{i2}}{p_{i2}/\pi_i}\right) = Y_i$$

$$\text{where } Y_i = \sum_{\text{Group } i} y_j$$

Therefore, $E(\hat{Y}_1) = Y$ . . . (1.3.2)

Thus $\hat{Y}_1$ is an unbiased estimate of the population total Y.

<u>Variance of $\hat{Y}_1$</u>:

$$V(\hat{Y}_1) = E_1V_2(\hat{Y}_1) + V_1E_2(\hat{Y}_1)$$

where V_2 and V_1 are variances for a given split and for all possible splits respectively.

$$V(\hat{Y}_1) = E_1V_2(\hat{Y}_1) \quad \text{because } V_1E_2(\hat{Y}_1) = 0$$

$$= \sum_{i=1}^{n/2} E_1V_2 \left[\frac{1}{2}\left(\frac{y_{i1}}{p_{i1}/\pi_i} + \frac{y_{i2}}{p_{i2}/\pi_i} \right) \right]$$

$$= \frac{1}{2}\sum_{i=1}^{n/2} E_1V_2 \left(\frac{y_{i1}}{p_{i1}/\pi_i} \right)$$

$$= \frac{1}{2}\sum_{i=1}^{n/2} \frac{N_i(N_i - 1)}{N(N-1)} \sum_{j<j'}^{N} p_jp_{j'}\left(\frac{y_j}{p_j} - \frac{y_{j'}}{p_{j'}} \right)^2$$

$$= \frac{1}{2} \frac{\sum_{i=1}^{n/2} N_i^2 - N}{N(N-1)} \sum_{j=1}^{N} p_j \left(\frac{y_j}{p_j} - Y \right)^2 \quad . . (1.3.3)$$

<u>Case 1</u>: $V(\hat{Y}_1)$ is minimum when all N_i's are equal. Therefore, when N is a multiple of $n/2$ setting $N_1 = N_2 = . . = N_{n/2} = 2N/n$, it follows from (1.3.3) that

$$V(\hat{Y}_1) = \frac{N - n/2}{n(N-1)} \sum_{j=1}^{N} p_j \left(\frac{y_j}{p_j} - Y \right)^2 \quad . . . (1.3.4)$$

Case 2:- Let N be not a multiple of n/2 and set $N = \frac{n}{2} \cdot P' + k'$, where P' and k' are integers and $k' < n/2$, then taking

$$N_1 = N_2 = \ldots = N_{k'} = P' + 1 \; ; \; N_{k'+1} = N_{k'+2} = \ldots N_{n/2} = P'$$

from (1.3.3) it is seen that

$$V(\hat{Y}_1) = \frac{(N - \frac{n}{2} + k')(N - k')}{nN(N-1)} \sum_{j=1}^{N} p_j \left(\frac{y_j}{p_j} - Y \right)^2 \quad \ldots \quad (1.3.5)$$

When n is odd, split up the population at random into(n-1)/2 groups and then draw 3 units from one and two units from the rest of the groups independently with probability proportional to p_t's and with replacement.

The estimator of the population total in this case is

$$\hat{Y}_1' = \sum_{i=1}^{(n-3)/2} \frac{1}{2} \left(\frac{y_{i1}}{p_{i1}/\pi_i} + \frac{y_{i2}}{p_{i2}/\pi_i} \right) + \frac{1}{3} \sum_{k=1}^{3} \frac{y_{jk}}{p_{jk}/\pi_j} \quad \ldots \quad (1.3.6)$$

where j denotes the group from which three units are selected. It can easily be seen that

$$E(\hat{Y}_1') = Y \quad \ldots \quad (1.3.7)$$

and

$$V(\hat{Y}_1') = \left[\frac{1}{2} \sum_{i=1}^{(n-3)/2} \frac{N_i(N_i - 1)}{N(N-1)} + \frac{1}{3} \frac{N_j(N_j - 1)}{N(N-1)} \right] \cdot \left\{ \sum_{t=1}^{N} p_t \left(\frac{y_t}{p_t} - Y \right)^2 \right\} \quad \ldots (1.3.8)$$

Case 1: - Let N be a multiple of (n-1)/2 and set

$$N_1 = N_2 = \ldots N_{\frac{n-1}{2}} = \frac{2N}{n-1} \; ; \quad \text{then}$$

$$V(\hat{Y}_1') = \frac{(N - \frac{n-1}{2})(n - \frac{5}{3})}{(n-1)^2 (N-1)} \sum_{t=1}^{N} p_t \left(\frac{y_t}{p_t} - Y \right)^2 \quad . . (1.3.9)$$

<u>Case 2:</u> - Again, if $N = \frac{n-1}{2} P'' + k''$, $\; 0 \quad k'' \quad \frac{n-1}{2}$

and $N_1 = N_2 = ... = N_{k''} = P'' + 1$; $N_{k''+1} = . . = N_{\frac{n-1}{2}} = P''$.

Then assuming $N_j = P'' + 1$, from (1.3.8), it follows that

$$V(\hat{Y}'_1) = \frac{P''^2 (\frac{n-1}{4} - \frac{1}{6}) + P'' (k'' - \frac{1}{6} - \frac{n-1}{4})}{N(N-1)} \cdot$$

$$\sum_{t=1}^{N} p_t \left(\frac{y_t}{p_t} - Y \right)^2 \quad . . . . (1.3.10)$$

Further, if $N_j = P''$, then

$$V(\hat{Y}'_1) = \frac{1}{N(N-1)} \left[(\frac{n-1}{4} - \frac{1}{6}) P'' (P'' - 1) + P'' k'' \right] .$$

$$\sum_{t=1}^{N} p_t \left(\frac{y_t}{p_t} - Y \right)^2 \quad . . . . \quad (1.3.11)$$

It is easy to verify that the variance given in (1.3.10) is less than that given in (1.3.11). Hence selecting three units from a group containing $(P'' + 1)$ units is preferable in this case i.e. when N is not a multiple of $(n-1)/2$.

<u>Estimation of variances:</u>

<u>Case 1:</u> - Let n be even. It is clear that

$$E \; \frac{1}{2} \left(\frac{y_{i1}^2}{p_{i1}^2/\pi_i} + \frac{y_{i2}^2}{p_{i2}^2/\pi_i} \right) = E \left(\frac{y_{i1}^2}{p_{i1}^2/\pi_i} \right) = E_1 E_2 \left(\frac{y_{i1}^2}{p_{i1}^2/\pi_i} \right)$$

where E_2 and E_1 are expectations for a given split and for all possible splits respectively. Hence

$$E \frac{1}{2} \left(\frac{y_{i1}^2}{p_{i1}^2 / \pi_i} + \frac{y_{i2}^2}{p_{i2}^2 / \pi_i} \right) = E_1 \sum_{t=1}^{N_i} \frac{y_t^2}{p_t} = \frac{N_i}{N} \sum_{t=1}^{N} \frac{y_t^2}{p_t}$$

Therefore,

$$E \sum_{i=1}^{n/2} \frac{1}{2} \left(\frac{y_{i1}^2}{p_{i1}^2/\pi_i} + \frac{y_{i2}^2}{p_{i2}^2/\pi_i} \right) = \sum_{t=1}^{N} \frac{y_t^2}{p_t} \qquad \ldots (1.3.12)$$

But, $E \left[\hat{Y}_1^2 - v(\hat{Y}_1) \right] = Y^2$

where $E \left[v(\hat{Y}_1) \right] = V(\hat{Y}_1)$, therefore,

$$\sum_{t=1}^{N} \frac{y_t^2}{p_t} - Y^2 = E \left[\sum_{i=1}^{n/2} \frac{1}{2} \left(\frac{y_{i1}^2}{p_{i1}^2/\pi_i} + \frac{y_{i2}^2}{p_{i2}^2 / \pi_i} \right) - \hat{Y}_1^2 + v(\hat{Y}_1) \right] \qquad \ldots (1.3.13)$$

From (1.3.3), it is clear that

$$V(\hat{Y}_1) = A \sum_{t=1}^{N} p_t \left(\frac{y_t}{p_t} - Y \right)^2 \quad \text{where } A = \frac{1}{2} \, \frac{\sum_{i=1}^{n/2} N_i^2 - N}{N(N-1)}$$

It follows from (1.3.13) that

$$Ev(\hat{Y}_1) = V(\hat{Y}_1) = A \, E \left[\sum_{i=1}^{n/2} \frac{1}{2} \left(\frac{y_{i1}^2}{p_{i1}^2/\pi_i} + \frac{y_{i2}^2}{p_{i2}^2/\pi_i} \right) - \hat{Y}_1^2 + v(\hat{Y}_1) \right].$$

Therefore,

$$v(\hat{Y}_1) = \frac{A}{1-A} \left[\sum_{i=1}^{n/2} \frac{\pi_i}{2} \left(\frac{y_{i1}^2}{p_{i1}^2} + \frac{y_{i2}^2}{p_{i2}^2} \right) - \hat{Y}_1^2 \right] \quad \ldots \quad (1.3.14)$$

<u>Case 2:</u> - Let n be odd. It can easily be shown as in the previous case, that

$$E \left[\sum_{i=1}^{(n-3)/2} \frac{1}{2} \left(\frac{y_{i1}^2}{p_{i1}^2/\pi_i} + \frac{y_{i2}^2}{p_{i2}^2/\pi_i} \right) + \frac{1}{3} \sum_{k=1}^{3} \frac{y_{jk}^2}{p_{jk}^2/\pi_j} \right]$$

$$= \sum_{t=1}^{N} \frac{y_t^2}{p_t} \quad \ldots \quad (1.3.15)$$

Therefore,

$$\sum_{t=1}^{N} \frac{y_t^2}{p_t} - Y^2 = E \left[\sum_{i=1}^{(n-3)/2} \frac{\pi_i}{2} \left(\frac{y_{i1}^2}{p_{i1}^2} + \frac{y_{i2}^2}{p_{i2}^2} \right) \right.$$

$$\left. + \frac{\pi_j}{3} \sum_{k=1}^{3} \frac{y_{jk}^2}{p_{jk}^2} - \hat{Y'}_1^2 + v(\hat{Y'}_1) \right]$$

From (1.3.8), $V(\hat{Y'}_1)$ is given by

$$V(\hat{Y'}_1) = B \sum_{t=1}^{N} p_t \left(\frac{y_t}{p_t} - Y \right)^2 \quad \text{where } B = \frac{1}{2} \sum_{i=1}^{\frac{n-3}{2}} \frac{N_i(N_i-1)}{N(N-1)} + \frac{1}{3} \frac{N_j(N_j-1)}{N(N-1)}$$

Hence,

$$v(\hat{Y}'_1) = \frac{B}{1-B} \Big[\sum_{i=1}^{\frac{n-3}{2}} \frac{\pi_i}{2} \left(\frac{y_{i1}^2}{p_{i1}^2} + \frac{y_{i2}^2}{p_{i2}^2} \right)$$

$$+ \frac{\pi_j}{3} \sum_{k=1}^{3} \frac{y_{jk}^2}{p_{jk}^2} - \hat{Y}_1^2 \Big] \qquad \ldots \quad (1.3.16)$$

1.4. The selection scheme considered in the previous subsection provides a sample whose effective size is less than or equal to n. Since from each group only two units are selected by p.p.s. with replacement scheme, the probability of repetition of units is quite small. But still it does furnish samples with repeated units and therefore it appears that the estimates considered in this case cannot be more efficient than those when the design always provides samples whose effective size equals n. So it will be worthwhile to consider the following without replacement scheme.

Another alternative to the R.H.C. scheme : -

Let n be even. Select n units from the population as follows:

(a) Split the population at random into n/2 groups of sizes $N_1, N_2, \ldots, N_{n/2}$ so that $N_1 + N_2 + \ldots + N_{n/2} = N$.

(b) Draw a sample of size 2 with probability proportional to p_t's and without replacement from each of these groups independently.

An Estimate of Y : - An estimate of the population total Y is

$$\hat{Y}_2 = \sum_{i=1}^{n/2} \frac{1}{2} (t_{i1} + t_{i2}) \quad \ldots . (1.4.1)$$

$$\left.\begin{aligned} \text{where} \quad t_{i1} &= \frac{y_{i1}}{p_{i1}/\pi_i} \\ \text{and} \quad t_{i2} &= y_{i1} + \frac{y_{i2}}{p_{i2}/\pi_i} \left(1 - \frac{p_{i1}}{\pi_i} \right) \end{aligned}\right] \quad \ldots (1.4.2)$$

$y_{i1}, y_{i2}, p_{i1}, p_{i2}$ and π_i have got the same meanings as in the previous subsection.

It is easily seen that

$$E(\hat{Y}_2) = E_1 E_2 (\hat{Y}_2) = E_1 (\sum_{i=1}^{n/2} Y_i) = Y \quad . . (1.4.3)$$

where E_1 and E_2 are same as in subsection (1.3). Thus, $\hat{Y}_2$ is an unbiased estimate of Y.

<u>Variance of $\hat{Y}_2$</u>:- To derive an expression for the variance of $\hat{Y}_2$, the following lemma will be used.

<u>Lemma</u>:- If $\bar{x}_m^{(1)}$, $\bar{x}_m^{(2)}$ and $\bar{x}_m^{(3)}$ are the sample means of $x^{(1)}$, $x^{(2)}$ and $x^{(3)}$ characteristic values of the units of a simple random sample of m units drawn from a finite population of N units, then

$$E \frac{\bar{x}_m^{(1)} \bar{x}_m^{(2)}}{\bar{x}_m^{(3)}} \doteq \frac{\bar{X}^{(1)} \bar{X}^{(2)}}{\bar{X}^{(3)}} + \frac{N-m}{Nm(N-1)} \Big[\frac{1}{\bar{X}^{(3)}} \sum_{j=1}^{N} x_j^{(1)} x_j^{(2)}$$

$$- \frac{\bar{X}^{(2)}}{(\bar{X}^{(3)})^2} \sum_{j=1}^{N} x_j^{(1)} x_j^{(3)} - \frac{\bar{X}^{(1)}}{(\bar{X}^{(3)})^2} \sum_{j=1}^{N} x_j^{(2)} x_j^{(3)}$$

$$+ \frac{\bar{X}^{(1)} \bar{X}^{(2)}}{(\bar{X}^{(3)})^3} \sum_{j=1}^{N} (x_j^{(3)})^2 \Big] \quad \ldots (1.4.4)$$

to the first order of approximation

where $E(\bar{x}_m^{(i)}) = \bar{X}^{(i)}$, $i = 1, 2, 3.$

Proof: - Suppose that

$$x_j^{(i)} = \bar{X}^{(i)} + \epsilon_j^{(i)} \qquad \ldots (1.4.5)$$

for $i = 1, 2$ and 3 and $j = 1, 2, \ldots N$

such that $E(\epsilon_j^{(i)}) = 0$. From (1.4.5) it is clear that

$$\bar{x}_m^{(i)} = \bar{X}^{(i)} + \epsilon_m^{(i)} \quad \text{for } i = 1, 2, 3.$$

Now,

$$\frac{\bar{x}_m^{(1)} \cdot \bar{x}_m^{(2)}}{\bar{x}_m^{(3)}} = \frac{(\bar{X}^{(1)} + \bar{\epsilon}_m^{(1)})(\bar{X}^{(2)} + \bar{\epsilon}_m^{(2)})}{(\bar{X}^{(3)} + \bar{\epsilon}_m^{(3)})}$$

$$= \frac{\bar{X}^{(1)} \cdot \bar{X}^{(2)}}{\bar{X}^{(3)}} \left(1 + \frac{\bar{\epsilon}_m^{(1)}}{\bar{X}^{(1)}}\right)\left(1 + \frac{\bar{\epsilon}_m^{(2)}}{\bar{X}^{(2)}}\right).$$

$$\left(1 + \frac{\bar{\epsilon}_m^{(3)}}{\bar{X}^{(3)}}\right)^{-1}$$

Assuming that $\left|\frac{\bar{\epsilon}_m^{(3)}}{\bar{X}^{(3)}}\right| < 1$ and expanding the last

factor by Binomial expansion, we get

$$\frac{\bar{x}_m^{(1)} \cdot \bar{x}_m^{(2)}}{\bar{x}_m^{(3)}} = \frac{\bar{X}^{(1)} \cdot \bar{X}^{(2)}}{\bar{X}^{(3)}} \cdot \left[1 + \frac{\bar{\epsilon}_m^{(1)}}{\bar{X}^{(1)}} + \frac{\bar{\epsilon}_m^{(2)}}{\bar{X}^{(2)}} - \frac{E\bar{\epsilon}_m^{(3)}}{\bar{X}^{(3)}} \right.$$

$$+ \frac{\bar{\epsilon}_m^{(1)} \; \bar{\epsilon}_m^{(2)}}{\bar{X}^{(1)} \; \bar{X}^{(2)}} - \frac{\bar{\epsilon}_m^{(1)} \; \bar{\epsilon}_m^{(3)}}{\bar{X}^{(1)} \; \bar{X}^{(3)}} - \frac{\bar{\epsilon}_m^{(2)} \; \bar{\epsilon}_m^{(3)}}{\bar{X}^{(2)} \; \bar{X}^{(3)}}$$

$$\left. + \frac{(\bar{\epsilon}_m^{(3)})^2}{(\bar{X}^{(3)})^2} + \cdots\cdots \right]$$

Therefore,

$$E\left(\frac{\bar{x}_m^{(1)} \cdot \bar{x}_m^{(2)}}{\bar{x}_m^{(3)}} \right) \doteq \frac{\bar{X}^{(1)} \cdot \bar{X}^{(2)}}{\bar{X}^{(3)}} \cdot E\left[1 + \frac{\bar{\epsilon}_m^{(1)} \; \bar{\epsilon}_m^{(2)}}{\bar{X}^{(1)} \; \bar{X}^{(2)}} \right.$$

$$\left. - \frac{\bar{\epsilon}_m^{(1)} \; \bar{\epsilon}_m^{(3)}}{\bar{X}^{(1)} \; \bar{X}^{(3)}} - \frac{\bar{\epsilon}_m^{(2)} \; \bar{\epsilon}_m^{(3)}}{\bar{X}^{(2)} \; \bar{X}^{(3)}} + \frac{(\bar{\epsilon}_m^{(3)})^2}{(\bar{X}^{(3)})^2} \right]$$

neglecting the terms of order $\frac{1}{m^\gamma}$, where $\gamma \geq 2$

Now, $E(\bar{\epsilon}_m^{(i)} \cdot \bar{\epsilon}_m^{(i')}) = \frac{N-m}{Nm} \cdot \frac{1}{N-1} \sum_{j=1}^{N} \epsilon_j^{(i)} \epsilon_j^{(i')}$ for $i < i'$

$$= \frac{N-m}{Nm} \cdot \frac{N}{N-1} \cdot \mu_{11}^{(i\,i')}$$

where $\mu_{11}^{(i\,i')} = \frac{1}{N} \sum_{j=1}^{N} (x_j^{(i)} - \bar{X}^{(i)})(x_j^{(i')} - \bar{X}^{(i')})$

and $E(\bar{\epsilon}_m^{(3)})^2 = \frac{N-m}{Nm} \cdot \frac{N}{N-1} \mu_2^{(3)}$

where $\mu_2^{(3)} = \frac{1}{N} \left[\sum_{j=1}^{N} (x_j^{(3)})^2 - (\overline{X}^{(3)})^2 \right]$

Therefore,

$$E\left(\frac{\overline{x}_m^{(1)} \cdot \overline{x}_m^{(2)}}{\overline{x}_m^{(3)}} \right) \doteq \frac{\overline{X}^{(1)} \overline{X}^{(2)}}{\overline{X}^{(3)}} \left[1 + \frac{N-m}{Nm} \cdot \frac{N}{N-1} \left\{ \frac{\mu_{11}^{(12)}}{\overline{X}^{(1)} \overline{X}^{(2)}} - \frac{\mu_{11}^{(13)}}{\overline{X}^{(1)} \overline{X}^{(3)}} - \frac{\mu_{11}^{(23)}}{\overline{X}^{(2)} \overline{X}^{(3)}} + \frac{(\mu_2^{(3)})^2}{(\overline{X}^{(3)})^2} \right\} \right]$$

$$= \frac{\overline{X}^{(1)} \overline{X}^{(2)}}{\overline{X}^{(3)}} \left[1 + \frac{N-m}{Nm} \frac{N}{N-1} \left\{ \frac{\sum_{j=1}^{N} x_j^{(1)} x_j^{(2)}}{\overline{X}^{(1)} \overline{X}^{(2)}} - \frac{\sum_{j=1}^{N} x_j^{(1)} x_j^{(3)}}{\overline{X}^{(1)} \overline{X}^{(3)}} - \frac{\sum_{j=1}^{N} x_j^{(2)} x_j^{(3)}}{\overline{X}^{(2)} \overline{X}^{(3)}} + \frac{\sum_{j=1}^{N} (x_j^{(3)})^2}{(\overline{X}^{(3)})^2} \right\} \right]$$

$$= \frac{\overline{X}^{(1)} \cdot \overline{X}^{(2)}}{\overline{X}^{(3)}} + \frac{N-m}{Nm} \cdot \frac{N}{N-1} \left[\frac{1}{\overline{X}^{(3)}} \sum_{j=1}^{N} x_j^{(1)} x_j^{(2)} - \frac{\overline{X}^{(2)}}{(\overline{X}^{(3)})^2} \sum_{j=1}^{N} x_j^{(1)} x_j^{(3)} - \frac{\overline{X}^{(1)}}{(\overline{X}^{(3)})^2} \sum x_j^{(2)} x_j^{(3)} + \frac{\overline{X}^{(1)} \cdot \overline{X}^{(2)}}{(\overline{X}^{(3)})^3} \sum_{j=1}^{N} (x_j^{(3)})^2 \right]$$

Thus, the lemma is proved.

Now, consider the variance of $\hat{Y}_2$, viz.

$$V(\hat{Y}_2) = E_1V_2(\hat{Y}_2) + V_1E_2(\hat{Y}_2)$$

where V_2 is the variance for a given split and V_1 is the variance for all possible splits of the population. Hence,

$$V(\hat{Y}_2) = E_1V_2(\hat{Y}_2)$$

$$= \sum_{i=1}^{n/2} E_1V_2 \left[\frac{1}{2}(t_{i1} + t_{i2}) \right]$$

where t_{i1} and t_{i2} are given by (1.4.2)

$$V(\hat{Y}_2) = \sum_{i=1}^{n/2} E_1 \frac{1}{4} \left[\sum_{j<j'}^{N} \frac{p_{ij}\, p_{ij'}}{\pi_i^2} \left(2 - \frac{p_{ij} + p_{ij'}}{\pi_i} \right) \left(\frac{y_{ij}}{p_{ij}/\pi_i} - \frac{y_{ij'}}{p_{ij'}/\pi_i} \right)^2 \right]$$

$$= \frac{1}{2} \sum_{i=1}^{n/2} E_1 \sum_{j<j'}^{N_i} p_{ij}\, p_{ij'} \left(\frac{y_{ij}}{p_{ij}} - \frac{y_{ij'}}{p_{ij'}} \right)^2$$

$$- \frac{1}{4} \sum_{i=1}^{n/2} E_1 \frac{1}{\pi_i} \left[\sum_{j<j'}^{N_i} p_{ij}\, p_{ij'} (p_{ij} + p_{ij'}) \left(\frac{y_{ij}}{p_{ij}} - \frac{y_{ij'}}{p_{ij'}} \right)^2 \right]$$

After a little simplification, we get

$$V(\hat{Y}_2) = \frac{1}{2} \sum_{i=1}^{n/2} E_1 \sum_{j<j'}^{N_i} p_{ij}\, p_{ij'} \left(\frac{y_{ij}}{p_{ij}} - \frac{y_{ij'}}{p_{ij'}} \right)^2$$

$$-\frac{1}{4} \sum_{i=1}^{n/2} E_1 \left[\frac{1}{\pi_i} \left\{ \pi_i \sum_{j=1}^{N_i} y_{ij}^2 + \left(\sum_{j=1}^{N_i} p_{ij}^2 \right)\left(\sum_{j=1}^{N_i} \frac{y_{ij}^2}{p_{ij}} \right) - 2Y \sum_{j=1}^{N_i} p_{ij}\, y_{ij} \right\} \right]$$

$$= \frac{1}{2} \sum_{i=1}^{(n/2)} \frac{N_i(N_i-1)}{N(N-1)} \sum_{j<j'}^{N} p_j\, p_{j'} \left(\frac{y_j}{p_j} - \frac{y_{j'}}{p_{j'}} \right)^2 - \frac{1}{4} \sum_{i=1}^{(n/2)} \frac{N_i}{N} \sum_{j=1}^{N} y_j^2$$

$$-\frac{1}{4} \sum_{i=1}^{n/2} \left[E_1 \frac{\left(\sum_{j=1}^{N_i} p_{ij}^2 \right)\left(\sum_{j=1}^{N_i} \frac{y_{ij}^2}{p_{ij}} \right)}{\left(\sum_{j=1}^{N_i} p_{ij} \right)} \right.$$

$$\left. - 2\, E_1 \frac{\left(\sum_{j=1}^{N_i} y_{ij} \right)\left(\sum_{j=1}^{N_i} p_{ij}\, y_{ij} \right)}{\left(\sum_{j=1}^{N_i} p_{ij} \right)} \right]$$

Now, since each random group can be considered as a simple random sample from the population, it follows from the above lemma, that

$$E \; \frac{\frac{1}{N_i}\left(\sum_{j=1}^{N_i} p_{ij}^2 \right) \frac{1}{N_i}\left(\sum_{j=1}^{N_i} \frac{y_{ij}^2}{p_{ij}} \right)}{\frac{1}{N_i} \sum_{j=1}^{N_i} p_{ij}} \doteq \frac{1}{N} \sum_{j=1}^{N} p_j^2 \sum_{j=1}^{N} \frac{y_j^2}{p_j} +$$

.. continued

$$+ \frac{N - N_i}{N N_i} \cdot \frac{N}{N-1} \Big[\sum_{j=1}^{N} y_j^2 p_j - (\sum_{j=1}^{N} y_j^2/p_j)(\sum_{j=1}^{N} p_j^3) - (\sum_{j=1}^{N} p_j^2)(\sum_{j=1}^{N} y_j^2)$$

$$+ (\sum_{j=1}^{N} p_j^2)^2 (\sum_{j=1}^{N} y_j^2/p_j) \Big]$$

and

$$E \frac{\frac{1}{N_i}(\sum_{j=1}^{N_i} y_{ij}) \frac{1}{N_i}(\sum_{j=1}^{N_i} y_{ij} p_{ij})}{\frac{1}{N_i}\sum_{j=1}^{N_i} p_{ij}} \doteq \frac{1}{N} Y \sum_{j=1}^{N} p_j y_j$$

$$+ \frac{N - N_i}{N N_i} \cdot \frac{N}{N-1} \Big[\sum_{j=1}^{N} y_j^2 p_j$$

$$-(\sum_{j=1}^{N} p_j y_j)^2 - Y \sum_{j=1}^{N} p_j^2 y_j + Y (\sum_{j=1}^{N} p_j y_j)(\sum_{j=1}^{N} p_j^2) \Big]$$

Therefore,

$$V(\hat{Y}_2) \doteq \frac{1}{2} \sum_{i=1}^{n/2} \frac{N_i(N_i-1)}{N(N-1)} \sum_{j<j'}^{N} p_j p_{j'} \left(\frac{y_j}{p_j} - \frac{y_{j'}}{p_{j'}} \right)^2$$

$$- \frac{1}{4} \sum_{i=1}^{n/2} \Big[\frac{N_i}{N} \sum_{j=1}^{N} y_j^2 + \frac{N_i}{N} \sum_{j=1}^{N} p_j^2 \sum_{j=1}^{N} \frac{y_j^2}{p_j} - \frac{2N_i}{N} Y \sum_{j=1}^{N} p_j y_j$$

$$+ \frac{N_i - N_i}{N-1} \Big\{ \sum_{j=1}^{N} p_j y_j^2 - \sum_{j=1}^{N} \frac{y_j^2}{p_j} \sum_{j=1}^{N} p_j^3 - \sum_{j=1}^{N} p_j^2 \sum_{j=1}^{N} y_j^2$$

$$+ (\sum_{j=1}^{N} p_j^2)^2 \sum_{j=1}^{N} y_j^2/p_j - 2 \sum_{j=1}^{N} p_j y_j^2 + 2(\sum_{j=1}^{N} p_j y_j)^2 + 2Y \sum_{j=1}^{N} p_j^2 y_j$$

$$- 2Y \sum_{j=1}^{N} p_j^2 \sum_{j=1}^{N} p_j y_j \Big\} \Big]$$

It is easily seen that

$$\sum_{j<j'}^{N} p_j p_{j'}(p_j+p_{j'})\left(\frac{y_j}{p_j} - \frac{y_{j'}}{p_{j'}}\right)^2 = \sum_{j=1}^{N} y_j^2 + \sum_{j=1}^{N} p_j^2 \sum_{j=1}^{N} \frac{y_j^2}{p_j} - 2Y\sum_{j=1}^{N} p p_j y_j$$

and

$$\sum_{j<j'}^{N} p_j p_{j'}(p_j+p_{j'})^2\left(\frac{y_j}{p_j} - \frac{y_{j'}}{p_{j'}}\right)^2 = \sum_{j=1}^{N} \frac{y_j^2}{p_j} \sum_{j=1}^{N} p_j^3 + \sum_{j=1}^{N} p_j y_j^2$$

$$+2\left(\sum_{j=1}^{N} p_j^2\right)\left(\sum_{j=1}^{N} y_j^2\right) - 2\left(\sum_{j=1}^{N} p_j y_j\right)^2 - 2Y\sum_{j=1}^{N} p_j^2 y_j$$

Therefore,

$$V(\hat{Y}_2) \doteq \frac{1}{2}\sum_{i=1}^{n/2} \frac{N_i(N_i-1)}{N(N-1)} \sum_{j<j'}^{N} p_j p_{j'}\left(\frac{y_j}{p_j} - \frac{y_{j'}}{p_{j'}}\right)^2$$

$$-\frac{1}{4}\sum_{i=1}^{n/2}\left(\frac{N_i}{N} + \frac{N-N_i}{(N-1)}\sum_{j=1}^{N} p_j^2\right)\sum_{j<j'}^{N} p_j p_{j'}(p_j+p_{j'})\left(\frac{y_j}{p_j} - \frac{y_{j'}}{p_{j'}}\right)^2$$

$$+\frac{1}{4}\sum_{i=1}^{n/2} \frac{N-N_i}{N-1} \sum_{j<j'}^{N} p_j p_{j'}(p_j+p_{j'})^2\left(\frac{y_j}{p_j} - \frac{y_{j'}}{p_{j'}}\right)^2 \quad ..(1.4.7)$$

<u>Case 1:</u> - Assuming N to be a multiple of $n/2$ and taking $N_1 = N_2 = \dots = N_{n/2} = 2N/n$, from (1.4.7) it follows that

$$V(\hat{Y}_2) = \frac{(N-\frac{n}{2})}{n(N-1)} \sum_{j<j'}^{N} p_j p_{j'}\left(\frac{y_j}{p_j} - \frac{y_{j'}}{p_{j'}}\right)^2$$

$$-\frac{1}{4}\left(1 + \frac{N}{N-1}\cdot\frac{n-2}{2}\sum_{j=1}^{N} p_j^2\right)\sum_{j<j'}^{N} p_j p_{j'}(p_j+p_{j'})\left(\frac{y_j}{p_j} - \frac{y_{j'}}{p_{j'}}\right)^2$$

$$+\frac{1}{4}\cdot\frac{N}{N-1}\cdot\frac{n-2}{2}\sum_{j<j'}^{N} p_j p_{j'}(p_j+p_{j'})^2\left(\frac{y_j}{p_j} - \frac{y_{j'}}{p_{j'}}\right)^2 \quad ..(1.4.8)$$

Case 2:- Let $N = \frac{n}{2} P' + k'$, where P' and k' are integers and $0 < k' < n/2$. Now taking

$N_1 = N_2 = \ldots\ldots = N_{k'} = P'+1$ s ; $N_{k+1} = N'_{k+2} = \ldots = N_{n/2} = P'$

from (1.4.7), it follows that

$$V(Y_2) = \frac{(N - \frac{n}{2} + k')(N - k')}{nN(N-1)} \sum_{j<j'}^{N} p_j p_{j'} \left(\frac{y_j}{p_j} - \frac{y_{j'}}{p_{j'}} \right)^2$$

$$- \frac{1}{4} \left(1 + \frac{N}{N-1} \cdot \frac{n-2}{2} \cdot \Sigma p_j^2 \right) \sum_{j<j'}^{N} p_j p_{j'} (p_j + p_{j'}) \left(\frac{y_j}{p_j} - \frac{y_{j'}}{p_{j'}} \right)^2$$

$$+ \frac{1}{4} \frac{N}{N-1} \cdot \frac{n-2}{2} \sum_{j<j'}^{N} p_j p_{j'} (p_j + p_{j'})^2 \left(\frac{y_j}{p_j} - \frac{y_{j'}}{p_{j'}} \right)^2$$

. . .(1.4.9)

Estimator of the variance:

Since $V(\hat{Y}_2) = E_1 V_2 (\hat{Y}_2)$

$$= \sum_{i=1}^{n/2} E_1 V_2 \frac{1}{2} (t_{i1} + t_{i2})$$

$$= \sum_{i=1}^{n/2} E_1 E_2 v_2 \left[\frac{1}{2} (t_{i1} + t_{i2}) \right]$$

where $E_2 v_2 \frac{1}{2}(t_{i1} + t_{i2}) = V_2 \frac{1}{2}(t_{i1} + t_{i2})$

so that $$V(\hat{Y}_2) = E \sum_{i=1}^{n/2} v_2 \left[\frac{1}{2} (t_{i1} + t_{i2}) \right]$$

$$= E \sum_{i=1}^{n/2} \frac{1}{4} \left(1 - \frac{p_{i1}}{\pi_i}\right) \left(1 - \frac{p_{i2}}{\pi_i}\right) \left(\frac{y_{i1}}{p_{i1}/\pi_i} - \frac{y_{i2}}{p_{i2}/\pi_i} \right)^2$$

Therefore, if $v(\hat{Y}_2)$ is the estimate of the variance of $\hat{Y}_2$, then

$$v(\hat{Y}_2) = \frac{1}{4} \sum_{i=1}^{n/2} (\pi_i - p_{i1}) (\pi_i - p_{i2}) \left(\frac{y_{i1}}{p_{i1}} - \frac{y_{i2}}{p_{i2}} \right)^2 \qquad \ldots (1.4,10)$$

When n is odd, the corresponding sampling procedure can be given as in the previous subsection(1.3). Since estimate of the population total and its variance can easily be developed on the same lines in this case as in the previous one, no further elaboration is intended.

Since the unordered Des Raj estimate (cf Murthy[3]) is more efficient than the ordered one, an estimate of Y with this scheme which is more efficient than $\hat{Y}_2$ can be obtained as follows:

Set

$$\hat{Y}_3 = \sum_{i=1}^{n/2} \frac{1}{\left(2 - \frac{p_{i1}}{\pi_i} - \frac{p_{i2}}{\pi_i}\right)} \left[\left(1 - \frac{p_{i2}}{\pi_i}\right) \frac{y_{i1}}{p_{i1}/\pi_i} + \left(1 - \frac{p_{i1}}{\pi_i}\right) \frac{y_{i2}}{p_{i2}/\pi_i} \right]$$

$$\hat{Y}_3 = \sum_{i=1}^{n/2} \frac{\pi_i}{(2\pi_i - p_{i1} - p_{i2})} \left[(\pi_i - p_{i2}) \frac{y_{i1}}{p_{i1}} + (\pi_i - p_{i1}) \frac{y_{i2}}{p_{i2}} \right]$$

In view of the selection scheme it is clear that $E(\hat{Y}_3) = Y$.

The variance of $\hat{Y}_3$ is

$$V(\hat{Y}_3) = \sum_{i=1}^{n/2} E_1 \sum_{j<j'}^{N_i} p_{ij}\, p_{ij'} \frac{\pi_i - p_{ij} - p_{ij'}}{2\pi_i - p_{ij} - p_{ij'}} \left(\frac{y_{ij}}{p_{ij}} - \frac{y_{ij'}}{p_{ij'}} \right)^2 \quad \ldots . (1.4.11)$$

and an estimate of the $V(\hat{Y}_3)$ is given by

$$v(\hat{Y}_3) = \sum_{i=1}^{n/2} \frac{\pi_i(\pi_i - p_{i1})(\pi_i - p_{i2})(\pi_i - p_{i1} - p_{i2})}{(2\pi_i - p_{i1} - p_{i2})^2} \left(\frac{y_{i1}}{p_{i1}} - \frac{y_{i2}}{p_{i2}} \right)^2$$

1.5. Comparison of estimates:-

In this subsection we consider the following estimates of Y for purposes of comparison:

(i) The standard estimate in case of varying probability with replacement :

$$\hat{Y}' = \frac{1}{n} \sum_{i=1}^{n} y_i/p_i$$

(ii) The Rao-Hartley-Cochran estimate:

$$\hat{Y} = \sum_{i=1}^{n} y_i/(p_i/\pi_i)$$

(iii) The estimate proposed in subsection (1.3):

$$\hat{Y}_1 = \sum_{i=1}^{n/2} \frac{\pi_i}{2} \left(\frac{y_{i1}}{p_{i1}} + \frac{y_{i2}}{p_{i2}} \right)$$

(iv) The estimate proposed in subsection (1.4):

$$\hat{Y}_2 = \sum_{i=1}^{n/2} \frac{\pi_i}{2} \left\{ \frac{y_{i1}}{p_{i1}} \left(1 + \frac{p_{i1}}{\pi_i}\right) + \frac{y_{i2}}{p_{i2}} \left(1 - \frac{p_{i1}}{\pi_i}\right) \right\}$$

For simplicity it is assumed that n is even in what follows. It is known that (cf [1])

$$V(\hat{Y}') \geqslant V(\hat{Y}) \qquad \ldots \quad (1.5.1)$$

Further, from (1.3.3), (1.4.7), (1.4.6) and (1.4.11) it is clear that

$$V(\hat{Y}_1) \geqslant V(\hat{Y}_2) \geqslant V(\hat{Y}_3) \qquad \ldots \quad (1.5.2)$$

where $\hat{Y}_3$ is the improved estimate proposed in the subsection (1.4). Since,

$$V(\hat{Y}') = \frac{1}{n} \left[\sum_{t=1}^{N} \frac{y_t^2}{p_t} - Y^2 \right] \ldots \quad (1.5.3)$$

therefore, from (1.5.3), (1.3.5) and (1.2.4) it is clear that

$$V(\hat{Y}') \geqslant V(\hat{Y}_1) \geqslant V(\hat{Y}) \qquad \ldots \quad (1.5.4)$$

holds true always.

This leads to the conclusion that the proposed estimate $\hat{Y}_1$ is better than the standard estimate in case of varyingeprobability with replacement but is worse than the R.H.C. estimate. Also, amongst the three proposed estimates $\hat{Y}_3$ is best but it cannot be compared with other estimates under consideration, because it's variance is not known in comparable form.

Comparison between $\hat{Y}_2$ and $\hat{Y}$:

Suppose that $N = nR + k$ where R and k are integers and $0 \leq k < n$, then from (1.2.4)

$$V(\hat{Y}) = \frac{(N-n+k)(N-k)}{nN(N-1)} \sum_{j<j'}^{N} p_j p_{j'} \left(\frac{y_j}{p_j} - \frac{y_{j'}}{p_{j'}}\right)^2 \quad \ldots (1.5.5)$$

Now assuming that $N = \frac{n}{2} P' + k'$

$k' = k - n/2 \quad$ if $k \geq n/2$

and $\quad k' = k \quad$ if $k < n/2$.

Thus,

if $k \geq n/2$, then $\dfrac{(N - \frac{n}{2} + k')(N-k')}{nN(N-1)} = \dfrac{(N-n+k)(N-k) + \frac{n}{2}(N-n+k)}{nN(N-1)}$

and if $k < n/2$ then $\dfrac{(N-\frac{n}{2}+k')(N-k')}{nN(N-1)} = \dfrac{(N-n+k)(N-k) + \frac{n}{2}(N-k)}{nN(N-1)}$

Therefore, if $k \geq n/2$ from (1.4.9) it follows that

$$V(\hat{Y}_2) = V(\hat{Y}) + \frac{1}{2}\frac{(N-n+k)}{N(N-1)} \sum_{j<j'}^{N} p_j p_{j'} \left(\frac{y_j}{p_j} - \frac{y_{j'}}{p_{j'}}\right)^2$$

$$\pm \frac{1}{4}\left\{1 + \frac{N}{(N-1)} \quad \frac{n-2}{2} \sum_{j=1}^{N} p_j^2\right\} \sum_{j<j'}^{N} p_j p_{j'}(p_j + p_{j'}) \cdot \left(\frac{y_j}{p_j} - \frac{y_{j'}}{p_{j'}}\right)^2$$

$$+\frac{1}{4}\cdot\frac{N}{N-1}\cdot\frac{n-2}{2}\sum_{j<j'}^{N} p_j\, p_{j'}\,(p_j+p_{j'})^2\left(\frac{y_j}{p_j}-\frac{y_{j'}}{p_{j'}}\right)^2$$

$$\ldots \quad (1.5.6)$$

and if $k < n/2$, then

$$V(\hat{Y}_2) = V(\hat{Y}) + \frac{1}{2}\cdot\frac{(N-k)}{N(N-1)}\cdot\sum_{j<j'}^{N} p_j\, p_{j'}\left(\frac{y_j}{p_j}-\frac{y_{j'}}{p_{j'}}\right)^2$$

$$-\frac{1}{4}\left\{1+\frac{N}{N-1}\;\frac{n-2}{2}\sum_{j=1}^{N} p_j^2\right\}\sum_{j<j'}^{N} p_j\, p_{j'}\,(p_j+p_{j'})\left(\frac{y_j}{p_j}-\frac{y_{j'}}{p_{j'}}\right)^2$$

$$+\frac{1}{4}\cdot\frac{N}{N-1}\cdot\frac{n-2}{2}\sum_{j<j'}^{N} p_j\, p_{j'}(p_j+p_{j'})^2\left(\frac{y_j}{p_j}-\frac{y_{j'}}{p_{j'}}\right)^2$$

$$\ldots \quad (1.5.7)$$

Hence, if $k \geq n/2$, then

$$V(\hat{Y}_2) - V(\hat{Y}) = \frac{1}{2}\cdot\frac{N-n+k}{N(N-1)}\cdot\sum_{j<j'}^{N} p_j\, p_{j'}\left(\frac{y_j}{p_j}-\frac{y_{j'}}{p_{j'}}\right)^2$$

$$-\frac{1}{4}\sum_{j<j'}^{N} p_j\, p_{j'}(p_j+p_{j'})\left(\frac{y_j}{p_j}-\frac{y_{j'}}{p_{j'}}\right)^2$$

$$-\frac{N}{(N-1)}\cdot\frac{n-2}{8}\sum_{j=1}^{N} p_j^2\sum_{j<j'}^{N} p_j\, p_{j'}(p_j+p_{j'})\left(\frac{y_j}{p_j}-\frac{y_{j'}}{p_{j'}}\right)^2$$

$$+\frac{N}{N-1}\cdot\frac{n-2}{8}\sum_{j<j'}^{N} p_j\, p_{j'}\,(p_j+p_{j'})^2\left(\frac{y_j}{p_j}-\frac{y_{j'}}{p_{j'}}\right)^2$$

Now since,

$$\sum_{j<j'}^{N} p_j\, p_{j'}\,(p_j+p_{j'})\left(\frac{y_j}{p_j}-\frac{y_{j'}}{p_{j'}}\right)^2 = \sum_{j=1}^{N} p_j^2\cdot\sum_{j=1}^{N} p_j\left(\frac{y_j}{p_j}-Y\right)^2$$

$$+\sum_{j=1}^{N} p_j^2\left(\frac{y_j}{p_j}-Y\right)^2$$

and $\sum_{j<j'}^{N} p_j p_{j'} \left(\frac{y_j}{p_j} - \frac{y_{j'}}{p_{j'}} \right)^2 = \sum_{j=1}^{N} p_j \left(\frac{y_j}{p_j} - Y \right)^2$

Therefore,

$$V(\hat{Y}_2) - V(\hat{Y}) = \left(\frac{N - n + k}{2N(N-1)} - \frac{1}{8} \sum_{j=1}^{N} p_j^2 \right) \sum_{j=1}^{N} p_j \left(\frac{y_j}{p_j} - Y \right)^2$$

$$- \frac{N}{N-1} \frac{n-2}{8} \sum_{j=1}^{N} p_j^2 \sum_{j<j'}^{N} p_j p_{j'} (p_j + p_{j'}) \left(\frac{y_j}{p_j} - \frac{y_{j'}}{p_{j'}} \right)^2 - \frac{1}{8} \sum_{j=1}^{N} p_j^2 \left(\frac{y_j}{p_j} - Y \right)^2$$

$$- \frac{n-2}{8} \sum_{j<j'}^{N} \left[p_j p_{j'} (p_j + p_{j'}) \left(\frac{y_j}{p_j} - \frac{y_{j'}}{p_{j'}} \right)^2 \left\{ \frac{1}{n-2} - \frac{N}{N-1} (p_j + p_{j'}) \right\} \right]$$

It is now clear that the r.h.s. is definitely negative if

$$\left.\begin{aligned} &\sum_{j=1}^{N} p_j^2 \geqslant \frac{4(N - n + k)}{N(N-1)} \\ &\text{and} \\ &\frac{1}{n-2} \geqslant \frac{N}{N-1} (p_j + p_{j'}) \text{ for every } j \text{ and } j' \end{aligned}\right\} \quad \ldots (1.5.8)$$

Even if for some cases the inequality $\frac{1}{n-2} < \frac{N}{N-1} (p_j + p_{j'})$ holds good, the r.h.s. can be expected to be negative when the population is quite skewed and therefore $V(\hat{Y}_2) < V(\hat{Y})$ in such cases.

Again if $k < n/2$ then $V(\hat{Y}_2) < V(\hat{Y})$ when

$$\left.\begin{aligned} &\sum_{j=1}^{N} p_j^2 \geqslant \frac{4(N-k)}{N(N-1)} \\ &\text{and} \\ &\frac{1}{n-2} \geqslant \frac{N}{(N-1)} (p_j + p_{j'}) \text{ for every } j \text{ and } j'. \end{aligned}\right\} \quad \ldots (1.5.9)$$

Section II

ON SOME UNEQUAL PROBABILITY SAMPLING TECHNIQUES WITH MISSING OBSERVATIONS

2.1. Introduction: - Sometimes, in sample surveys, inspite of all precautions, some observations might be missing due to some unfoSseen reasons. For example, in agricultural yield surveys, it is very common to meet with a situation when due to some practical difficulties, the crop in a selected field is harvested before an observation is made on it. It is, therefore, interesting to study the behaviour of the estimates considered in the last section in such a situation. In this section besides comparing the efficiencies of these estimates, the ratio estimate is compared with R.H.C. estimate with missing observations. It is found that in some situations the estimatesproposed in the previous section are better than the R.H.C. estimate in case of missing observations. Further, the ratio estimate is superior to the R.H.C. estimate in many cases particularly when ρ (the correlation coefficient between y and x) is very high.

2.2. Effect of missing observations on R.H.C. estimate

When no observation is missing the R.H.C. estimate is given by (1.2.1) i.e.

$$\hat{Y} = \sum_{i=1}^{n} \frac{y_i}{p_i/\pi_i}$$

where the suffixes 1,2,...,n denote the n units selected from the n groups separately.

Now, suppose that due to some unforseen causes γ observations are missing. An estimate corresponding to R.H.C. scheme based on (n - γ) observations is

$$\hat{Y}_b = \sum_{i=1}^{n-\gamma} \frac{y_i}{p_i/\pi_i} \qquad \ldots (2.2.1)$$

where the suffixes 1,2,...,(n - γ) denote the (n - γ) available selected units from n groups separately. Now

$$E(\hat{Y}_b) = E_1 E_2 (\hat{Y}_b)$$

where E_2 and E_1 have got the same meanings as in sub-section 1.2.

$$E_2(\hat{Y}_b) = \sum_{i=1}^{n-\gamma} E_2 \left(\frac{y_i}{p_i/\pi_i} \right)$$

$$= \sum_{i=1}^{n-\gamma} Y_i \quad \text{where} \quad Y_i = \sum_{j=1}^{N_i} y_j$$

$$= Y - \sum_{\gamma}{}' Y_i$$

where $\sum_{\gamma}'$ is the summation over all those γ groups in which observations are missing. Therefore,

$$E(\hat{Y}_b) = Y - E_1 \left(\sum_{\gamma}{}' Y_i \right)$$

$$= Y - \sum_{\gamma}{}' \frac{N_i}{N} \quad Y$$

$$= \frac{N - \sum_{\gamma}' N_i}{N} \quad Y \qquad \ldots (2.2.2)$$

Thus, we see that $\hat{Y}_b$ is a biased estimate of Y, and the statistic

$$\hat{Y}_u = \frac{N}{N - \Sigma'_{\gamma} N_i} \hat{Y}_b \qquad \ldots (2.2.3)$$

is an unbiased estimate of Y.

Case 1 : - Let N be a multiple of n. Then, taking

$$N_1 = N_2 = \ldots = N_n = N/n ,$$

it follows from (2.2.3) that

$$\hat{Y}_u = \frac{n}{n - \gamma} \hat{Y}_b \qquad \ldots (2.2.4)$$

Case 2:- Suppose that N = nR+k, where R and k are integers and k < n. Then, taking

$$N_1 = N_2 = \ldots = N_k = R + 1 ; \quad N_{k+1} = N_{k+2} = \ldots = N_n = R$$

and supposing that γ_1 observations are missing from the groups of size R + 1 and γ_2 observation are missing from the group of size R so that $\gamma_1 + \gamma_2 = \gamma$, we get

$$\hat{Y}_u = \frac{N}{N - \gamma_1 (R + 1) - \gamma_2 R} \hat{Y}_b$$

$$= \frac{n}{(n-\gamma) + \frac{1}{N}(\gamma k - n\gamma_1)} \hat{Y}_b \quad \ldots (2.2.5)$$

Variance of $\hat{Y}_u$ - The variance of $\hat{Y}_b$ is

$$V(\hat{Y}_b) = E_1 V_2(\hat{Y}_b) + V_1 E_2 (\hat{Y}_b).$$

where V_2 and V_1 denote the variances for a given split and for all possible splits of the population into groups of sizes $N_1, N_2, \ldots, N_n$ respectively. Thus,

$$V(\hat{Y}_b) = \sum_{n-\gamma} E_1 V_2 \left(\frac{y_i}{p_i / \pi_i} \right) + V_1 (Y - \sum_{\gamma}' Y_i)$$

$$= \sum_{n-\gamma} \frac{N_i(N_i - 1)}{N(N-1)} \sum_{t=1}^{N} p_t \left(\frac{y_t}{p_t} - Y \right)^2 + V_1 (\sum_{\gamma}' Y_i) \quad \ldots . \quad (2.2.6)$$

where $\sum_{n-\gamma}$ is the summation over those $(n-\gamma)$ groups in which observations are available.

Now, $V_1 (\sum_{\gamma}' Y_i) = V_1 \left[(\sum_{\gamma}' N_i) (\bar{Y}_{\sum_{\gamma}' N_i}) \right]$

where $\bar{Y}_{\sum_{\gamma}' N_i}$ is the pooled average of y's falling in the groups containing missing observations. Since each group can be considered as a random sample from the population, $\bar{Y}_{\sum_{\gamma}' N_i}$ can be considered as the mean of a random sample of size $(\sum_{\gamma}' N_i)$ from the population.

Hence

$$V_1 (\sum_{\gamma}' Y_i) = (\sum_{\gamma}' N_i)^2 \left(\frac{1}{\sum_{\gamma}' N_i} - \frac{1}{N} \right) S_y^2$$

where

$$(N-1) S_y^2 = \sum_{i=1}^{N} (y_i - \bar{y}_N)^2$$

Therefore,

$$V(\hat{Y}_b) = \sum_{n-\gamma} \frac{N_i(N_i-1)}{N(N-1)} \sum_{t=1}^{N} p_t \left(\frac{y_t}{p_t} - Y\right)^2 + \left(\sum_{\gamma}' N_i\right)^2 .$$

$$\left(\frac{1}{\sum_{\gamma}' N_i} - \frac{1}{N}\right) S_y^2 \quad \ldots (2.2.7)$$

and $$V(\hat{Y}_u) = \frac{N^2}{(N - \sum_{\gamma}' N_i)^2} \cdot V(\hat{Y}_b) \quad \ldots (2.2.8)$$

Case 1: - Let N be a multiple of n. Assuming $N_1 = N_2 = \ldots = N_n = N/n$, the equation (2.2.8) simplifies to

$$V(\hat{Y}_u) = \frac{N-n}{(n-\gamma)(N-1)} \left[\sum_{t=1}^{N} \frac{y_t^2}{p_t} - Y^2 \right] + \frac{\gamma N}{n-\gamma} S_y^2$$

$$\ldots (2.2.9)$$

Case 2: - Let $N = nR + k$, where R and k are integers and $0 < k < n$. Taking $N_1 = N_2 = \ldots = N_k = R + 1$; $N_{k+1} = \ldots = N_n = R$ and supposing that γ_1 and γ_2 are number of observations missing from the groups of size R + 1 and R respectively, we get

$$\frac{N}{N - \sum_{\gamma}' N_i} = \frac{n}{(n-\gamma) + \frac{1}{N}(\gamma k - \gamma_1 n)} ,$$

$$\sum_{n-\gamma} \frac{N_i(N_i-1)}{N(N-1)} = \frac{(N-k)(N-k-n)(n-\gamma) + 2n(N-k)(k-\gamma_1)}{n^2 N(N-1)}$$

and

$$\left(\sum_{\gamma} N_i\right)^2 \left(\frac{1}{\sum_{\gamma}' N_i} - \frac{1}{N}\right) = \frac{\gamma(N-k)+n\gamma_1}{n^2} \left\{(n-\gamma) + \frac{1}{N}(\gamma k - \gamma_1 n)\right\}$$

So that from (2.2.8), it follows that

$$V(\hat{Y}_u) = \frac{(N-k)(N-k-n)(n-\gamma)+2n(N-k)(k-\gamma_1)}{N(N-1)\left\{n-\gamma+\frac{1}{N}(\gamma k-\gamma_1 n)\right\}^2} \sum_{t=1}^{N} p_t\left(\frac{y_t}{p_t}-Y\right)^2$$

$$+ \frac{\gamma(N-k)+n\gamma_1}{(n-\gamma)+\frac{1}{N}(\gamma k-\gamma_1 n)} \cdot S_y^2 \quad \ldots . \quad (2.2.10)$$

It reduces to (2.2.9) when $k = 0$ and $\gamma_1 = 0$.

2.3. Effect of missing observations on $\hat{Y}_1$: -

To study the effects of missing observations on the estimate $\hat{Y}_1$, suppose that out of $n/2$ groups s are such that both the observations are missing from each, μ are such that only one is missing from each and the data are available from all the units selected from the remaining groups, so that on the whole $\gamma = 2s + u$ units of the sample are missed.

A biased estimate of the population total Y is

$$\hat{Y}_{b(1)} = \sum_{(\frac{n}{2}-s-u)} \frac{1}{2}\left(\frac{y_{i1}}{p_{i1}/\pi_i} + \frac{y_{i2}}{p_{i2}/\pi_i}\right) + \sum_u{}' \frac{y_j}{p_j/\pi_j}$$

$$\ldots \quad (2.3.1)$$

where $\sum_{(\frac{n}{2}-s-u)}$ denotes summation over the $(\frac{n}{2}-s-u)$ groups which provide both the observations and $\sum_u{}'$ is the summation over the u groups which provide only one observation each.

It is clear that

$$E(\hat{Y}_{b(1)}) = \frac{N - \Sigma''_{s} N_i}{N} . Y \qquad \ldots (2.3.2)$$

where the summation Σ''_{s} is taken over the s groups in each of which both the units selected are missing. Thus $\hat{Y}_{b(1)}$ is a biased estimate of Y.

From (2.3.2) it is clear that the estimate $\hat{Y}_{u(1)}$ given by

$$\hat{Y}_{u(1)} = \frac{N}{N - \Sigma''_{s} N_i} . \hat{Y}_{b(1)} \qquad \ldots (2.3.3)$$

is unbiased for Y.

Case 1:- Let N be a multiple of n/2. Set $N_1 = N_2 = \ldots = N_n$ in (2.3.3). Then, clearly

$$\hat{Y}_{u(1)} = \frac{n}{n - 2s} . \hat{Y}_{b(1)} \qquad \ldots (2.3.4)$$

Case 2:- Suppose that $N = \frac{n}{2} P' + k'$ where P' and k' are integers and $0 < k' < n/2$. Put

$$N_1 = N_2 = \ldots\ldots = N_{k'} = P'+1 \; ; \; N_{k'+1} = \ldots = N_{n/2} = P'.$$

Now assuming that of the s groups s_1 belong to the class of groups each of which contains $P' + 1$ units and s_2 belong to the other category of groups so that $s_1 + s_2 = s$, it is found that

$$\frac{N}{N - \Sigma''_{s} N_i} = \frac{n}{n - 2s \left(\frac{N - k'}{N}\right) - \frac{ns_1}{N}}$$

Thus (2.3.3) reduces to

$$\hat{Y}_{u(1)} = \frac{n}{n - 2s(1 - k'/N) - ns_1/N} \cdot \hat{Y}_{b(1)} \quad \ldots (2.3.5)$$

when $k' = 0$ and $s_1 = 0$, (2.3.5) reduces to (2.3.4)

Variance of the estimate $\hat{Y}_{u(1)}$: -

Variance of $\hat{Y}_{b(1)}$ is given by

$$V(\hat{Y}_{b(1)}) = E_1 V_2 (\hat{Y}_{b(1)}) + V_1 E_2 (\hat{Y}_{b(1)})$$

where E_1, V_1, E_2 and V_2 carry the same meanings as in subsection (1.3). Hence

$$V(\hat{Y}_{b(1)}) = \Big[\frac{1}{2} \sum_{(\frac{n}{2} - s - u)} \frac{N_i(N_i - 1)}{N(N-1)} + \Sigma' \frac{N_i(N_i - 1)}{N(N-1)} \Big] \sum_{t=1}^{N} p_t \Big(\frac{Y_t}{p_t} - Y \Big)^2 + \Big(\sum_s{}'' N_i \Big)^2 \Big(\frac{1}{\sum_s'' N_i} - \frac{1}{N} \Big) S_y^2 \quad \ldots (2.3.6)$$

and

$$V(\hat{Y}_{u(1)}) = \Big[\frac{N}{N - \sum_s'' N_i} \Big] V(\hat{Y}_{b(1)}) \quad \ldots (2.3.7)$$

Case 1: - Let N be a multiple of n/2. Putting $N_1 = N_2 \ldots = N_{n/2}$

we get, that $\frac{N}{N - \sum_s'' N_i} = \frac{n}{n - 2s} = \frac{n}{n - \gamma + u}$,

$$\sum_{(\frac{n}{2} - s - u)} \frac{N_i(N_i - 1)}{N(N-1)} = \frac{(n - \gamma - u)(2N - n)}{n^2 (N-1)},$$

$$\Sigma'_u \frac{N_i(N_i - 1)}{N(N-1)} = \frac{2u(2N - n)}{n^2 (N-1)}.$$

and $\left(\sum_{s}'' N_i\right)^2 \left(\frac{1}{\sum_{s}'' N_i} - \frac{1}{N}\right) = \frac{(\gamma-u)(n-\gamma+u)N}{n^2}$

so that $$V(\hat{Y}_{u(1)}) = \frac{(n-\gamma+3u)(N - \frac{n}{2})}{(n-\gamma+u)^2(N-1)} \sum_{t=1}^{N} p_t\left(\frac{y_t}{p_t} - Y\right)^2$$

$$+ \frac{(\gamma-u)N}{(n-\gamma+u)} S_y^2 \qquad \ldots (2.3.8)$$

When $\gamma = u$ i.e. not more than one observation is missing from a group (2.3.8) reduces to

$$V(\hat{Y}_{u(1)}) = \frac{(n+2\gamma)(N - \frac{n}{2})}{n^2(N-1)} \sum_{t=1}^{N} p_t\left(\frac{y_t}{p_t} - Y\right)^2$$

$$\ldots (2.3.9)$$

<u>Case 2:-</u> When $N = \frac{n}{2} P' + k'$ where P' and k' are integers and $0 < k' < n/2$, put $N_1 = N_2 = \ldots\ldots = N_{k'} = P'+1$ and $N_{k'+1} = \ldots = N_{n/2} = P'$.

Let $\left(\frac{n}{2} - s - u\right)_1$ = Number of groups of size $(P'+1)$ such that no observation is missing from them.

$\left(\frac{n}{2} - s - u\right)_2$ = Number of groups of size P' such that no observation is missing from them.

u_1 = Number of groups of size $(P'+1)$ with one missing observation.

u_2 = Number of groups of size P' with one missing observation.

Then,

$$\frac{N}{N - \sum_{s}'' N_i} = \frac{N}{N - (s P' + s_1)}$$

$$\sum_{(\frac{n}{2}-s-u)} \frac{N_i(N_i-1)}{N(N-1)} = \frac{1}{N(N-1)} \Big[(\frac{n}{2}-s-u)P'^2$$

$$+ P' \{(\frac{n}{2}-s-u)_1 - (\frac{n}{2}-s-u)_2\} \Big]$$

$$\sum_u{}' \frac{N_i(N_i-1)}{N(N-1)} = \frac{1}{N(N-1)} \Big[u P'^2 + P'(u_1-u_2) \Big]$$

and

$$(\sum_s{}'' N_i)^2 \left(\frac{1}{\sum_s^o N_i} - \frac{1}{N}\right) = (sP'+s_1) \Big[1 - \left(\frac{sP'+s_1}{N}\right) \Big]$$

Hence,

$$V(\hat{Y}_{u(1)}) = \frac{1}{\left(1 - \frac{sP'+s_1}{N}\right)^2} \Big[\frac{1}{N(N-1)} \Big\{ \frac{1}{2}\left(\frac{n}{2}-s+u\right)P'^2$$

$$+ \frac{1}{2}P'((\frac{n}{2}-s-u)_1 - (\frac{n}{2}-s-u)_2) + P'(u_1-u_2)\Big\} \sum_{t=1}^{N} p_t \left(\frac{y_t}{p_t} - Y\right)^2$$

$$+ (sP'+s_1)\left(1 - \frac{sP'+s_1}{N}\right) S_y^2 \Big] \quad . . \ (2.3.10)$$

It is easy to verify that when $k' = 0$ i.e. $(\frac{n}{2}-s-u)_1 = u_1 = s_1 = 0$, then (2.3.10) reduces to (2.3.9).

2.4. Effect of missing observations on $\hat{Y}_2$: -

Under the selection scheme considered in sub-section(1.4) proceeding on the same lines as before, we see that if $\gamma(=2s+u)$ observations are missing such that 2s observations are missing from s groups and u observations are missing from u groups, the statistic

$$\hat{Y}_{u(2)} = \frac{N}{N - \Sigma''_{s} N_i} \cdot \hat{Y}_{b(2)} \quad \ldots \quad (2.4.1)$$

is an unbiased estimate of Y, where

$$\hat{Y}_{b(2)} = \sum_{(\frac{n}{2}-s-u)} \frac{1}{2}\,(t_{i1} + t_{i2}) + \Sigma'_{u} \frac{y_j}{p_j/\pi_j} \,,$$

it is being assumed that $t_{i1} = \frac{y_{i1}}{p_{i1}/\pi_i}$ and

$t_{i2} = y_{i1} + \frac{y_{i2}}{p_{i2}/\pi_i}\,(1 - p_{i1}/\pi_i)$, Σ, Σ' and Σ'' have the same meanings as given in the previous subsection.

<u>Variance of $\hat{Y}_{u(2)}$</u> :- The variance of $\hat{Y}_{b(2)}$ is given by

$$V(\hat{Y}_{b(2)}) = \frac{1}{2} \sum_{(\frac{n}{2}-s-u)} \frac{N_i(N_i - 1)}{N(N-1)} \sum_{j<j'}^{N} p_j p_{j'} \left(\frac{y_j}{p_j} - \frac{y_{j'}}{p_{j'}}\right)^2$$

$$- \frac{1}{4} \sum_{(\frac{n}{2}-s-u)} \left(\frac{N_i}{N} + \frac{N - N_i}{N-1} \cdot \sum_{j=1}^{N} p_j^2\right)^2 .$$

$$\sum_{j<j'}^{N} p_j p_{j'} (p_j + p_{j'}) \left(\frac{y_j}{p_j} - \frac{y_{j'}}{p_{j'}}\right)^2$$

$$+ \frac{1}{4} \sum_{(\frac{n}{2}-s-u)} \frac{N - N_i}{N-1} \sum_{j<j'}^{N} p_j p_{j'} (p_j + p_{j'})^2 \left(\frac{y_j}{p_j} - \frac{y_{j'}}{p_{j'}}\right)^2$$

--- continued

$$+ \sum_{u}' \frac{N_i(N_i - 1)}{N(N-1)} \sum_{j<j'}^{N} p_j p_{j'} \left(\frac{y_j}{p_j} - \frac{y_{j'}}{p_{j'}} \right)^2$$

$$+ \left(\sum_{s}'' N_i \right)^2 \left(\frac{1}{\sum_{s}'' N_i} - \frac{1}{N} \right) S_y^2$$

$$= V(\hat{Y}_{b(1)}) - \frac{1}{4} \sum_{(\frac{n}{2} - s - u)} \left(\frac{N_i}{N} + \frac{N - N_i}{N-1} \sum_{j=1}^{N} p_j^2 \right).$$

$$\sum_{j<j'}^{N} p_j p_{j'} (p_j + p_{j'}) \left(\frac{y_j}{p_j} - \frac{y_{j'}}{p_{j'}} \right)^2 + \frac{1}{4} \sum_{(\frac{n}{2} - s - u)} \frac{N - N_i}{N-1} \cdot$$

$$\sum_{j<j'}^{N} p_j p_{j'} (p_j + p_{j'})^2 \left(\frac{y_j}{p_j} - \frac{y_{j'}}{p_{j'}} \right)^2 \quad \ldots (2.4.2)$$

and

$$V(\hat{Y}_{u(2)}) = \left(\frac{N}{N - \sum_{s}'' N_i} \right)^2 V(\hat{Y}_{b(2)}) \quad \ldots (2.4.3)$$

2.5. Comparison of efficiencies

In this subsection the efficiencies of the following estimates in case of missing observations are compared.

(i) Standard estimate in case of varying probability with replacement:

$$\hat{Y}'_{u'} = \frac{1}{n - \gamma} \sum_{i=1}^{n-\gamma} \frac{y_i}{p_i}$$

(ii) Estimate corresponding to Rao, Hartley and Cochran procedure (cf [2.2]):

$$\hat{Y}_u = \frac{N}{N - \Sigma'_{\gamma} N_i} \sum_{n-\gamma} \frac{y_i}{p_i / \pi_i}$$

(iii) The estimate proposed in subsection 1.3 (cf [2.3])

$$\hat{Y}_{u(1)} = \frac{N}{N - \Sigma''_{s} N_i} \left[\sum_{(\frac{n}{2} - s - u)} \frac{1}{2} \left(\frac{y_{i1}}{p_{i1}/\pi_i} + \frac{y_{i2}}{p_{i2}/\pi_i} \right) \right]$$

(iv) The estimate proposed in subsection 1.4 [cf (2.4)]

$$\hat{Y}_{u(2)} = \frac{N}{N - \Sigma''_{s} N_i} \left[\sum_{(\frac{n}{2} - s - u)} \frac{1}{2} (t_{i1} + t_{i2}) + \Sigma'_{u} \frac{y_j}{p_j / \pi_j} \right]$$

For purposes of simplicity the comparison of efficiencies is restricted to the case where N is a multiple of n which is assumed to be even. Now,

$$V(\hat{Y}'_u) = \frac{1}{n - \gamma} \sigma_z^2 \quad \ldots \quad (2.5.a)$$

where $\sigma_z^2 = \sum_{t=1}^{N} p_t \left(\frac{y_t}{p_t} - Y \right)^2$

$$V(\hat{Y}_u) = \frac{(N - n)}{(n - \gamma)(N-1)} \sigma_z^2 + \frac{\gamma N}{n - \gamma} S_y^2 \quad \ldots \quad (2.5.b)$$

$$V(\hat{Y}_{u(1)}) = \frac{(n - \gamma + 3u)}{(n - \gamma + u)^2} \frac{(N - \frac{n}{2})}{N - 1} \sigma_z^2 + \frac{\gamma - u}{(n - \gamma + u)} N S_y^2 \quad \ldots \quad (2.5.c)$$

and

$$V(\hat{Y}_{u(2)}) = V(\hat{Y}_{u(1)}) - \frac{n^2(n-\gamma-u)}{8(n-\gamma+u)^2} \Big[\left(\frac{2}{n} + \frac{N}{N-1} \frac{n-2}{n} \sum_{j=1}^{N} p_j^2 \right).$$

$$\sum_{j<j'}^{N} p_j\, p_{j'} (p_j + p_{j'}) \left(\frac{y_j}{p_j} - \frac{y_{j'}}{p_{j'}} \right)^2 - \frac{N}{N-1} \; \frac{n-2}{n} \, .$$

$$\sum_{j<j'}^{N} p_j\, p_{j'} (p_j + p_{j'})^2 \left(\frac{y_j}{p_j} - \frac{y_{j'}}{p_{j'}} \right)^2 \Big] \quad \ldots (2.5.d)$$

2.5.1. Comparison between $\hat{Y}'_u$ and $\hat{Y}_u$:-

Since sampling with varying probability is used primarily in the situation in which y_t's are approximately proportional to p_t's , a reasonable model, which is considered by Cochran, is

$$y_t = Yp_t + e_t$$

where e_t is independent of p_t in the probability sense and in the arrays in which p_t is fixed

$$E(e_t) = 0 \quad \text{and} \quad E(e_t^2) = ap_t^g \qquad (g > 0)$$

Unless otherwise stated all the subsequent comparisons will be made under this model. It is easy to verify that under this model

$$S_y^2 = R^2 S_x^2 + S_e^2$$

where $p_t = \dfrac{x_t}{X}$, $X = \sum_{t=1}^{N} x_t$, $R = \dfrac{Y}{X}$, $S_x^2 = \dfrac{1}{(N-1)} \sum_{t=1}^{N} (x_t - \bar{x}_N)^2$,

$$S_e^2 = \frac{1}{N-1} \sum_{t=1}^{N} e_t^2 \quad \text{and} \quad S_y^2 = \frac{1}{(N-1)} \sum_{t=1}^{N} (y_i - \bar{y}_N)^2$$

Now, $\sum_{t=1}^{N} e_t^2 = N\,E(e_t^2) = N\,E\,E(e_t^2 \mid p_t) = aNE(p_t^g)$

where the average is taken over all values of p_t's. Similarly

$$\sum_{t=1}^{N} \frac{e_t^2}{p_t} = aNE(p_t^{g-1}) = aN^2 \left[E(p_t^g) - \text{Cov.}(p_t, p_t^{g-1}) \right]$$

Now, from (2.5.a.) and (2.5.b) it follows that

$$V(\hat{Y}_u) - V(\hat{Y}') = \frac{\gamma N}{n-\gamma} S_y^2 - \frac{n-1}{(n-\gamma)(N-1)} \sum_{t=1}^{N} p_t \left(\frac{y_t}{p_t} - Y \right)^2$$

$$= \frac{\gamma N}{n-\gamma} (R^2 S_x^2 + S_e^2) - \frac{n-1}{(n-\gamma)(N-1)} \sum_{t=1}^{N} \frac{e_t^2}{p_t}$$

$$= \frac{\gamma N}{n-\gamma} \left[R^2 S_x^2 + \frac{aN}{(N-1)} E(p_t^g) \right]$$

$$- \frac{(n-1)(aN^2}{(n-\gamma)(N-1)} \left[E(p_t^g) - \text{Cov}(p_t, p_t^{g-1}) \right]$$

$$= \frac{a\gamma N^2 E(p_t^g)}{(n-\gamma)(N-1)} \left(\frac{R^2 S_x^2}{S_e^2} - \frac{n-\gamma-1}{\gamma} \right)$$

$$+ \frac{(n-1)\, aN^2}{(n-\gamma)(N-1)} \text{Cov.}\,(p_t, p_t^{g-1})$$

because $S_e^2 = \frac{aN}{(N-1)} E(p_t^g)$.

Now, if ρ is the correlation coefficient between y_t's and p_t's, then

$$S_e^2 = (1 - \rho^2)\, S_y^2$$

or
$$\frac{R^2 S_x^2}{S_e^2} = \frac{\rho^2}{1 - \rho^2}$$

Therefore,

$$V(\hat{Y}_u) - V(\hat{Y}'_u) = \frac{a\gamma N^2 E(p_t^g)}{(n-\gamma)(N-1)} \left[\frac{\rho^2}{1-\rho^2} - \frac{n-\gamma-1}{\gamma} \right.$$

$$\left. + \frac{(n-1)\, aN^2}{(n-\gamma)(N-1)} \operatorname{Cov}(p_t, p_t^{g-1}) \right]$$

$$= \frac{(n-1)\, aN^2}{(n-\gamma)(N-1)} \left[\frac{E(p_t^g)}{(1-\rho^2)} \left(\rho^2 - \frac{n-\gamma-1}{n-1}\right) \right.$$

$$\left. + \operatorname{Cov}(p_t, p_t^{g-1}) \right] \qquad \ldots . \; (2.5.1.1)$$

Now, from (2.5.1.1) it follows that since $\operatorname{Cov}(p_t, p_t^{g-1}) \gtrless 0$ according as $g \gtrless 1$

(i) if $g < 1$ and $\rho^2 < \frac{n-\gamma-1}{n-1}$ then definitely $V(\hat{Y}_u) < V(\hat{Y}'_u)$

(ii) if $g > 1$ and $\rho^2 > \frac{n-\gamma-1}{n-1}$ then definitely $V(\hat{Y}_u) > V(\hat{Y}'_u)$

(iii) if $g=1$ then $V(\hat{Y}_u) \gtrless V(\hat{Y}'_u)$ according as $\rho^2 \gtrless \frac{n-\gamma-1}{n-1}$.

The inequality $\rho^2 > \frac{n-\gamma-1}{n-1}$ may be satisfied if ρ is very high and $\frac{\gamma}{n-1}$ is not very small. For example, if $n = 11$, $\gamma = 2$ and $\rho = .9$ then the above inequality is satisfied. It is known that g usually lies between 1 and 2. Hence in most of the cases if

$\rho^2 > \frac{n-\gamma-1}{n-1}$ then $\hat{Y}'_u$ is superior to $\hat{Y}'_u$. On the other hand if $g < 1$ and $\rho^2 < \frac{n-\gamma-1}{n-1}$, then $\hat{Y}'_u$ is inferior to $\hat{Y}_u$. It is interesting to consider the case when $g = 2$. After some simplification it is seen that

$$V(\hat{Y}_u) - V(\hat{Y}'_u) = \frac{aN^2}{(n-\gamma)(N-1)} \left[\frac{\gamma}{1-\rho^2} \quad E(p_t^2) - \frac{n-1}{N^2} \right]$$

when $g = 2$

Thus $\quad V(\hat{Y}_u) \geqslant V(\hat{Y}'_u)$

if $\quad E(p_t^2) \geqslant \frac{n-1}{N^2} \quad \frac{(1-\rho^2)}{\gamma}$

or if $\quad N \sum_{t=1}^{N} p_t^2 \geqslant \frac{n-1}{\gamma} (1-\rho^2)$

This inequality can be expected to be holding true when the population is skew and ρ is quite large. Thus we see that the universal superiority of R.H.C. estimate over the estimate in the case of varying probability with replacement is lost due to missing observations.

2.5.2 Comparison between $\hat{Y}'_u$ and $\hat{Y}_{u(1)}$:

Supposing that not more than one observation is missing from any group, so that $r = u$, we get from (2.5.a) and (2.5.c),

$$V(\hat{Y}'_u) - V(\hat{Y}_{u(1)}) = \left[\frac{1}{n-\gamma} - \frac{(n+2\gamma)(N-\frac{n}{2})}{n^2(N-1)} \right] \sum_{t=1}^{N} p_t \left(\frac{y_t}{p_t} - Y \right)^2$$

Thus $V(\hat{Y}'_u) \gtrless V(\hat{Y}_{u(1)})$ according as

$$\frac{1}{n-\gamma} \gtrless \frac{(n+2\gamma)(N-\frac{n}{2})}{n^2(N-1)}$$

or $$\frac{n}{N} \gtrless \frac{2\gamma(n-2\gamma)}{n(n-2)+\gamma(n-2\gamma)}$$

Since observations are expected to be missing when the sample size is quite large(and consequently the sampling fraction is large unless N is very large), the inequality

$$\frac{n}{N} > \frac{2\gamma(n-2\gamma)}{n(n-2)+\gamma(n-2\gamma)}$$

is more likely to be true. Hence, in such situations, the proposed estimate $\hat{Y}_{u(1)}$ is superior to $\hat{Y}'_u$. For example, if $N = 100$, $n = 20$, $\gamma = 2$ then the above inequality is satisfied and consequently $\hat{Y}_{u(1)}$ is more precise than $\hat{Y}'_u$. However, if sampling fraction is small $\hat{Y}'_u$ may be expected to be better than $\hat{Y}_{u(1)}$.

2.5.3. Comparison between $\hat{Y}_u$ and $\hat{Y}_{u(1)}$: -

From equations (2.5.b) and (2.5.c), we have

$$V(\hat{Y}_u) - V(\hat{Y}_{u(1)}) = \left(\frac{\gamma}{n-\gamma} - \frac{\gamma-u}{n-\gamma+u}\right) NS_y^2$$

$$- \left(\frac{n-\gamma+3\mu}{(n-\gamma+u)^2} \cdot \frac{N-\frac{n}{2}}{N-1} - \frac{N-n}{(n-\gamma)(N-1)}\right)\sigma_z^2$$

$$= ANS_y^2 - B\sigma_z^2 \qquad \ldots \qquad (2.5.3.1)$$

where $A = \dfrac{nu}{(n-\gamma)(n-\gamma+u)}$

and $B = \dfrac{1}{(N-1)}\left\{\dfrac{u(n-\gamma-u)(N-\frac{n}{2})}{(n-\gamma)(n-\gamma+u)^2} + \dfrac{n}{2(n-\gamma)}\right\}$

under the same model as in subsection (2.5.1) we get

$$V(\hat{Y}_u) - V(\hat{Y}_{u(1)}) = AN(R^2S_x^2 + S_e^2) - BaN^2\{E(p_t^g) - Cov(p_t, p_t^{g-1})\}$$

$$= \frac{aAN^2 E(p_t^g)}{(N-1)} \left[\frac{\rho^2}{1-\rho^2} - (N-1)\frac{B}{A} - 1\right]$$

$$+ BaN^2 \, Cov(p_t, p_t^{g-1})$$

Now, after a little simplification we get that

$$(N-1)\frac{B}{A} - 1 = \left(\frac{N}{n} + \frac{n-\gamma}{2u}\right) \frac{(n-\gamma-u)}{(n-\gamma+u)}$$

Therefore,

$$V(\hat{Y}_u) - V(\hat{Y}_{u(1)}) = \frac{a n u\, N^2\, E(p_t^g)}{(n - \gamma)(n-\gamma+u)(N-1)} \left[\frac{\rho^2}{1-\rho^2} - \left(\frac{N}{n} + \frac{n-\gamma}{2u} \right) \frac{n-\gamma-u}{n-\gamma+u} \right] + \frac{aN^2}{(N-1)} \left\{ \frac{u(n-\gamma-u)(N - \frac{n}{2})}{(n-\gamma)(n-\gamma+u)^2} + \frac{n}{2(n-\gamma)} \right\} Cov(p_t, p_t^{g-1}) \quad \ldots (2.5.3.2)$$

Now, (i) if $g < 1$ and

$$\frac{\rho^2}{1-\rho^2} < \left(\frac{N}{n} + \frac{n-\gamma}{2u} \right) \frac{n-\gamma-u}{n-\gamma+u} \quad \ldots (2.5.3.3)$$

then definitely, $V(\hat{Y}_u) < V(\hat{Y}_{u(1)})$. The inequality (2.5.3.3) is easily satisfied when sampling fraction is small and ρ is not very high, e.g. for $\rho = .6$, $N = 100$, $n = 10$ and $\gamma = u = 2$ this inequality is satisfied and $\hat{Y}_u$ is better than $\hat{Y}_{u(1)}$.

(ii) if $g > 1$ and

$$\frac{\rho^2}{1-\rho^2} > \left(\frac{N}{n} + \frac{n-\gamma}{2u} \right) \frac{(n-\gamma-u)}{(n-\gamma+u)} \quad \ldots (2.5.3.4)$$

then definitely $V(\hat{Y}_u) \quad V(\hat{Y}_{u(1)})$. The inequality (2.5.3.4) may be expected to be satisfied in small populations with high sampling fraction and very high correlation coefficient ρ, e.g. when $\rho = .9$, $N = 50$, $n = 10$ and $\gamma = 2$ the inequality holds.

(iii) if g = 1, then $V(\hat{Y}_u) \gtrless V(\hat{Y}_{u(1)})$ according as

$$\frac{\rho^2}{1-\rho^2} \gtrless \left(\frac{N}{n} + \frac{n-\gamma}{2u} \right) \frac{n-\gamma-u}{n-\gamma+\gamma}$$

Further, if g = 2, assuming that r = u (i.e. not more than one observation is missing from any group), it follows from (2.5.3.1) that $V(\hat{Y}_u) \gtrless V(\hat{Y}_{u(1)})$ according as

$$N \sum_{t=1}^{N} p_t^2 \gtrless (1-\rho^2) \left[\frac{(n-2\gamma)(N-\frac{n}{2})}{n^2} + \frac{n}{2\gamma} \right] \quad \ldots (2.5.3.5)$$

It is clear from (2.5.3.5) that for g = 2, $\hat{Y}_{u(1)}$ may be expected to be better than $\hat{Y}_u$ if the population is skew, ρ is high, N, n and γ are such that the inequality

$$N \sum_{t=1}^{N} p_t^2 > (1-\rho^2) \left[\frac{(n-2\gamma)(N-\frac{n}{2})}{n^2} + \frac{n}{2\gamma} \right]$$

is satisfied, for example, if $N \sum_{t=1}^{N} p_t^2 = 4$, $\rho^2 = .6$, N = 100, n = 20, γ = 2 then $V(\hat{Y}_u) > V(\hat{Y}_{u(1)})$.

Further, it is clear from (2.5.3.1) that if u = 0 i.e. γ observations are missing from γ/2(= s) groups, then $\hat{Y}_u$ is definitely better than $\hat{Y}_{u(1)}$ without restriction of any model.

2.5.4. Comparison between $\hat{Y}_{u(2)}$ and $\hat{Y}'$:

Setting $\gamma = u$ in (2.5.d) we get

$$V(\hat{Y}_{u(2)}) = \left[\frac{(n+2\gamma)}{n^2} \cdot \frac{N-\frac{n}{2}}{N-1} - \frac{(n-2\gamma)}{8n} \sum_{j=1}^{N} p_j^2 \right] \sigma_z^2$$

$$- \frac{n-2\gamma}{8} \cdot \frac{N}{N-1} \left[\frac{n-2}{n} \sum_{j=1}^{N} p_j^2 \sum_{j<j'}^{N} p_j p_{j'} (p_j + p_{j'}) \left(\frac{y_j}{p_j} - \frac{y_{j'}}{p_{j'}} \right)^2 \right.$$

$$- \frac{n-2\gamma}{8n} \sum_{j=1}^{N} p_j^2 \left(\frac{y_j}{p_j} - Y \right)^2 - \left(\frac{(n-2)(n-2\gamma)}{8n} \right) \cdot \left[\sum_{j<j'}^{N} p_j p_{j'} \cdot \right.$$

$$(p_j + p_{j'}) \left(\frac{y_j}{p_j} - \frac{y_{j'}}{p_{j'}} \right)^2 \left\{ \frac{1}{n-2} - \frac{N}{N-1} (p_j + p_{j'}) \right\} \Big]$$

Since in most of the cases the last term on the r.h.s. is expected to be negative, we may write

$$V(\hat{Y}_{u(2)}) \leq \left[\frac{n+2\gamma}{n^2} \cdot \frac{N-\frac{n}{2}}{N-1} - \frac{n-2\gamma}{8n} \sum_{j=1}^{N} p_j^2 \right] \sigma_z^2 \qquad \ldots . \text{ (2.5.4.1)}$$

from (2.5.a) and (2.5.4.1), it follows that

$$V(\hat{Y}') - V(\hat{Y}_{u(2)}) \geq \left[\frac{1}{n-\gamma} - \frac{(n+2\gamma)(N-n/2)}{n^2(N-1)} + \frac{n-2\gamma}{8n} \sum_{j=1}^{N} p_j^2 \right] \sigma_z^2$$

Thus $\quad V(\hat{Y}') \geq V(\hat{Y}_{u(2)})$

$$\text{if} \quad \frac{n-2\gamma}{8n} \sum_{j=1}^{N} p_j^2 \geqslant \frac{(n+2\gamma)(N-\frac{n}{2})}{n^2(N-1)} - \frac{1}{(n-\gamma)}$$

$$\text{or if} \quad \frac{(N-1)}{4} \sum_{j=1}^{N} p_j^2 \geqslant \frac{2\gamma\,(N-\frac{n}{2})}{n(n-\gamma)} - \frac{n(n-2)}{(n-\gamma)(n-2\gamma)} \qquad \ldots . \ (2.5.4.2)$$

In highly skewed populations, where $\Sigma p_j^2 \geqslant \frac{4}{N-1}$ holds, the inequality (2.5.4.2) can be expected to be satisfied in which case $\hat{Y}_{u(2)}$ is superior to $\hat{Y}'_u$.

2.5.5. Comparison between $\hat{Y}_u$ and $\hat{Y}_{u(2)}$

From (2.5.4.1) and (2.5.b) it follows that when $\gamma = u$

$$V(\hat{Y}_u) - V(\hat{Y}_{u(2)}) \geqslant \frac{\gamma N}{n-\gamma} S_y^2 - \Big[\frac{(n+2\gamma)(N-\frac{n}{2})}{n^2(N-1)} - \frac{n-2\gamma}{8n} \sum_{j=1}^{N} p_j^2 - \frac{N-n}{(n-\gamma)(N-1)} \Big] \sigma_s^2$$

$$= AN\, S_y^2 - B\sigma_s^2 \qquad \ldots . \ (2.5.5.1)$$

where $A = \gamma/(n-\gamma)$ and

$$B = \frac{(n+2\gamma)(N-\frac{n}{2})}{n^2(N-1)} - \frac{n-2\gamma}{8n} \sum_{j=1}^{N} p_j^2 - \frac{N-n}{(n-\gamma)(N-1)}$$

Now, $B \gtrless 0$ according as

$$\frac{n+2\gamma}{n^2} \cdot \frac{N-\frac{n}{2}}{N-1} - \frac{N-n}{(n-\gamma)(N-1)} \gtrless \frac{n-2\gamma}{8n} \sum_{j=1}^{N} p_j^2$$

or according as

$$\frac{n^2}{(n-\gamma)(n-2\gamma)} + \frac{2\gamma(\frac{N}{2} - \frac{1}{2})}{(n-\gamma)} \gtrless \frac{(N-1)}{4} \sum_{j=1}^{N} p_j^2$$

$$\ldots\ldots(2.5.5.2)$$

Now, it follows from (2.5.5.2) that if the population is very skew and the sampling fraction is not very small then B may be less than or equal to zero in which case it is clear from (2.5.5.1) that $\hat{Y}_{u(2)}$ is more efficient than $\hat{Y}_u$.

When $B > 0$ i.e. $\frac{n^2}{(n-\gamma)(n-2\gamma)} + \frac{2\gamma(\frac{N}{n} - \frac{1}{2})}{(n-\gamma)} > \frac{(N-1)}{4} \sum_{j=1}^{N} p_j^2$

we consider the model discussed in subsection(2.5.1). The inequality (2.5.5.1) reduces to

$$V(\hat{Y}_u) - V(\hat{Y}_{u(2)}) \geqslant \frac{AaN^2}{N-1} E(p_t^g)\left[\frac{\rho^2}{1-\rho^2} - \left\{(N-1)\frac{B}{A} - 1\right\}\right]$$

$$+ BaN^2 \text{Cov}(p_t, p_t^{g-1}) \quad \ldots (2.5.5.3)$$

After a little simplification it may be verified that

$$(N-1)\frac{B}{A} - 1 = \frac{n-2\gamma}{n}\left(\frac{N}{n} - \frac{1}{2} + \frac{n}{2\gamma}\right) - \frac{(n-2\gamma)(n-\gamma)}{8n\gamma}(N-1)\sum_{j=1}^{N} p_j^2$$

If $\sum_{j=1}^{N} p_j^2 \geqslant \frac{4}{N-1}$, then

$$(N-1)\frac{B}{A} - 1 \leqslant \frac{n-2\gamma}{n}\left(\frac{N}{n} - \frac{1}{2} + \frac{n}{2\gamma}\right) - \frac{(n-2\gamma)(n-\gamma)}{2n\gamma}$$

$$= \frac{(n-2\gamma)N}{n^2}$$

Thus, from (2.5.5.3), it follows that

$$V(\hat{Y}_u) - V(\hat{Y}_{u(2)}) \geq \frac{AaN^2}{(N-1)} \; E(p_t^g) \left[\frac{\rho^2}{1-\rho^2} - \frac{(n-2\gamma)N}{n^2} \right]$$

$$+ BaN^2 \, Cov(p_t, p_t^{g-1}).$$

Since $B > 0$, for $g \geq 1$, $V(Y_u) \geq V(Y_{u(2)})$ if

$$\frac{\rho^2}{1-\rho^2} \geq \frac{(n-2\gamma)N}{n^2} \qquad \ldots (2.5.5.4)$$

This inequality may be expected to hold good, if ρ is very high and sampling fraction is also high. But at the same time when sampling fraction is high, B may be expected to be negative in which case $\hat{Y}_{u(2)}$ is superior to $\hat{Y}_u$. Thus mostly, a higher sampling fraction alongwith the skewness of the population and a high value of ρ ensures the superiority of $\hat{Y}_{u(2)}$ over $\hat{Y}_u$ under the model considered here.

2.6. Comparison of ratio estimate with the R.H.C. estimate

A comparison between R.H.C. estimate and the ratio estimate under simple random sampling is of some practical interest. Since the bias in $\hat{Y}_R$ is negligible when sample size is large the comparison under reference is restricted to large samples only. The variance of the ratio estimate, viz.,

$$\hat{Y}_R = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i} \left(\sum_{i=1}^{N} x_i \right) \quad \ldots \quad (2.6.1)$$

is given, to the first order of approximation, by

$$V(\hat{Y}_R) = \frac{N(N-n)}{n(N-1)} \sum_{i=1}^{N} (y_i - Rx_i)^2 \quad \ldots \quad (2.6.2)$$

where $R = \bar{y}_N / \bar{x}_N$. Putting $p_t = x_i / N\bar{x}_N$ (in (2.6.2), it follows that

$$V(\hat{Y}_R) = \frac{N(N-n)}{n(N-1)} \sum_{i=1}^{N} (y_i - Y p_i)^2 \quad \ldots \quad (2.6.3)$$

Also, the variance of the R.H.C. estimate is given by

$$V(\hat{Y}) = \frac{N-n}{n(N-1)} \sum_{i=1}^{N} \frac{1}{p_i} (y_i - Yp_i)^2 \quad \ldots \quad (2.6.4)$$

where N is a multiple of n and $N_1 = N_2 = \ldots = N_n = N/n$.

Under the model considered in subsection (2.5.1), (2.6.3) reduces to

$$V(\hat{Y}_R) = \frac{N(N-n)}{n(N-1)} \sum_{t=1}^{N} e_t^2 = \frac{N(N-n)}{n(N-1)} \, aNE(p_t^g) \quad \ldots \quad (2.6.5)$$

where the expectation is now taken over all values of p_t's.

Also, from (2.6.4)

$$V(\hat{Y}) = \frac{N-n}{n(N-1)} \, aNE(p_t^{g-1}) \quad \ldots \quad (2.6.6)$$

Therefore,

$$V(\hat{Y}_R) - V(\hat{Y}) = \frac{N-n}{n(N-1)} \, aN^2 \left[E(p_t^g) - \frac{1}{N} E(p_t^{g-1}) \right]$$

$$= \frac{N-n}{n(N-1)} \, aN^2 \, \mathrm{Cov}(p_t, p_t^{g-1})$$

Since $\mathrm{Cov}(p_t, p_t^{g-1}) \gtrless 0$ accordingly as $g \gtrless 1$

Therefore $V(\hat{Y}_R) \gtrless V(\hat{Y})$ according as $g \gtrless 1$

. . . (2.6.7)

Now, since g usually lies between 1 and 2, we may expect $\hat{Y}$ to be superior to $\hat{Y}_R$ in such cases. This corraborates the result already obtained independently by Avdhani and Sukhatme under a different finite population model (cf [6]).

When γ observations are missing, the ratio estimate based on (n-γ) observations is

$$\hat{Y}_{Rm} = \frac{\sum_{i=1}^{n-\gamma} y_i}{\sum_{i=1}^{n-\gamma} x_i} \left(\sum_{i=1}^{N} x_i \right) \quad \ldots (2.6.8)$$

and

$$V(\hat{Y}_{Rm}) = \frac{N[N-(n-\gamma)]}{(n-\gamma)(N-1)} \sum_{j=1}^{N} (y_j - Yp_j)^2 \quad \ldots (2.6.9)$$

Also from (2.2.9)

$$V(\hat{Y}_u) = \frac{N-n}{(n-\gamma)(N-1)} \sum_{j=1}^{N} \frac{1}{p_j} (y_j - Yp_j)^2 + \frac{\gamma N}{(n-\gamma)} S_y^2$$

. . . (2.6.10)

under the model considered in (2.5.1)

$$V(\hat{Y}_u) - V(\hat{Y}_{Rm}) = \frac{\gamma N}{n-\gamma} (R^2 S_x^2 + S_e^2) + \frac{N-n}{(n-\gamma)(N-1)} \sum_{t=1}^{N} e_t^2 / p_t$$

$$< \frac{N(N-n+\gamma)}{(n-\gamma)(N-1)} \sum_{t=1}^{N} e_t^2$$

After a little simplification we see that

$$V(\hat{Y}_u) - V(\hat{Y}_{Rm}) = \frac{\gamma N}{n-\gamma} R^2 S_x^2 - \frac{(N-n)aN^2}{(n-\gamma)(N-1)} \cdot Cov(p_t, p_t^{g-1})$$

where the covariance is taken over all p_t's. Now, since

$S_e^2 = a \frac{N}{N-1} E(p_t^g)$, therefore

$$V(\hat{Y}_u) - V(\hat{Y}_{Rm}) = \frac{\gamma N}{n-\gamma} R^2 S_x^2 - \frac{(N-n)N}{(n-\gamma)} S_e^2 \frac{Cov(p_t, p_t^{g-1})}{E(p_t^g)}$$

Thus $V(\hat{Y}_u) \geqslant V(\hat{Y}_{Rm})$ if

$$\frac{R^2 S_x^2}{S_e^2} \geqslant \frac{(N-n)Cov(p_t, p_t^{g-1})}{\gamma E(p_t^g)}$$

or if,
$$\frac{\rho^2}{1-\rho^2} \geqslant \frac{(N-n)}{\gamma} \cdot \frac{Cov(p_t, p_t^{g-1})}{E(p_t^g)} \quad \ldots (2.6.11)$$

When $g \leq 1$, this inequality always holds good. Thus, even in the case when some observations are missing the ratio estimate is superior to the R.H.C. estimate when $g \leqslant 1$. Moreover, for $\rho \rightarrow 1$, this inequality always holds. So, when g is slightly greater than unity and ρ is sufficiently large, it appears that the ratio estimate is superior to the R.H.C. estimate under the model considered here.

SUMMARY

When some sample observations are missing, the R.H.C. estimate of the population total, based on the available observations, is not unbiased. A corresponding unbiased estimate has been considered which shows a considerable increase in its variance. Two alternatives to the R.H.C. scheme are proposed and the corresponding estimation procedures have been developed. The two proposed estimates are compared with the unbiased estimate corresponding to the R.H.C. scheme with and without missing observations under the well known Cochran's finite population model. It is found that in some situations the proposed schemes provide better estimates than the R.H.C. scheme. The unbiased estimate of the population total under the R.H.C. scheme with missing observations is also compared with the corresponding ratio estimate as derived from a simple random sample. It is seen that the ratio estimate is more efficient than the R.H.C. estimate when $g \leq 1$ and also when g is slightly greater than unity and ρ is sufficiently high.

REFERENCES

1. Rao, J.N.K., Hartley, H.O. and Cochran, W.G. (1962) "On a simple procedure of unequal probability sampling without replacement", J. Roy. Statis. Soc. B, pp. 482-491.

2. Des Raj (1956) " Some estimators in sampling with varying probabilities without replacement", J. Amer. Statis, Ass. Vol. 51 pp 269-284.

3. Murthy, M.N. (1957) " Ordered and unordered estimators in sampling without replacement", Sankhya Vol. 18, pp 379-390.

4. Cochran, W.G. (1953) "Sampling techniques", John Wiley and Sons.

5. Sukhatme, P.V. (1954) " Sampling theories of surveys with applications", Ind. Soc. of Agril. Statis.

6. Avadhani, M.S. and Sukhatme, B.V. (1967) To appear