ON SIMPLE RANDOM SAMPLING WITH REPLACEMENT

By P. K. PATHAK

Indian Statistical Institute

SUMMARY. In simple random sampling with replacement, Basu (1958), and Des Raj and Khamis (1958), showed that for estimating the population mean, the average of distinct units is more efficient than the overall sample mean. In this paper, a detailed treatment of the above problem is given, and the exact expression for the variance of above estimator is derived. The relative efficiency of the above estimator with other estimators is also considered. An improved estimator of the population variance is obtained. Finally, a comparison between the two simple random sampling schemes (with and without replacement) is made.

1. Introduction

We index the N population units as $1, 2, \ldots, N$, and let Y_j be some real-valued characteristic (in which we are interested) of the j-th population unit.* Here we consider the problem of estimating the population mean

$$\bar{Y} = N^{-1} \Sigma \; Y_j$$

and the population variance $\quad \sigma^2 = N^{-1} \Sigma \; (Y_j - \bar{Y})^2.$

For simplicity we refer population units by capital letters and sample units by small letters, e.g., u_i and y_i will denote the unit index and the variate value respectively associated with the i-th sample unit.

2. Estimation of $\bar{Y}$

In simple random sampling (with replacement), Basu (1958) considered two estimators of the population mean

(i) $\bar{y} = 1/n \; \Sigma \; y_i \equiv$ average of n sample units;

(ii) $\bar{y}_\nu = 1/\nu \; \Sigma \; y_{(i)} \equiv$ average of ν distinct units observed in the sample.

If we record the sample of observation as

$$S = (x_1, x_2, \ldots, x_n),$$

where $x_i = (y_i, u_i)$; and if ν be the number of distinct units observed in the sample, Basu (1958) showed that the 'order-statistic' (sample units arranged in ascending order of their unit-indices)

$$T = [x_{(1)}, x_{(2)}, \ldots, x_{(\nu)}]$$

(where $x_{(i)} = (y_{(i)}, u_{(i)})$, and $y_{(i)}$ is the variate value of the sample unit with unit index $u_{(i)}$) forms a sufficient statistic, and therefore, for any convex (downwards) loss function

$$E(\bar{y} \mid T) = E(y_1 \mid T) = \bar{y}_\nu \qquad \ldots \quad (2.1)$$

has uniformly smaller risk than $\bar{y}$.

An exact expression for variance of $\bar{y}_\nu$ is given below.

Variance of $\bar{y}_\nu$. We have

$$V(\bar{y}_\nu) = E\{V(\bar{y}_\nu \mid \nu)\} = E\left(\frac{1}{\nu} - \frac{1}{N}\right) S^2 \qquad \ldots \quad (2.2)$$

where $\quad S^2 = [N/(N-1)]\sigma^2.$

* j runs from 1 to N; i from 1 to n; and (i) from (1) to (ν).

Since $$E\left(\frac{1}{\nu}\right) = \frac{1^{n-1}+2^{n-1}+\ldots+N^{n-1}}{N^n}, \quad \text{(Pathak, 1961)}$$

$$V(\bar{y}_\nu) = \frac{1^{n-1}+2^{n-1}+\ldots+(N-1)^{n-1}}{N^n} S^2. \quad \ldots \quad (2.3)$$

For large samples, it is rather cumbersome to compute $V(\bar{y}_\nu)$. An approximate expression for $V(\bar{y}_\nu)$ valid for terms up to order N^{-2} is given by

$$V(\bar{y}_\nu) \doteq \left[\frac{1}{n} - \frac{1}{2N} + \frac{(n-1)}{12N^2}\right] S^2. \quad \ldots \quad (2.4)$$

3. Admissibility properties of certain estimators of $\bar{Y}$

Let Γ denote a certain class of estimators of $\bar{Y}$. For a given loss function, let $R(t)$ represent the risk (or expected loss) associated with the estimator t of $\bar{Y}$.

Of the two estimators t_1 and t_2 of $\bar{Y}$, t_1 will be said to be uniformly better than t_2 if, for a given loss function,

$$R(t_1) \leqslant R(t_2) \quad \ldots \quad (3.1)$$

holds for all possible values of $(Y_1, Y_2, \ldots, Y_N)$ with strict sign of inequality holding for at least one $(Y_1, Y_2 \ldots, Y_N)$.

An estimator t belonging to Γ is said to be admissible in Γ if there exists no estimator in Γ which is better than t.

Now we consider the problem of finding admissible estimators of $\bar{Y}$. As the 'order-statistic' T is sufficient, we have to restrict ourselves to functions of T only. Moreover, the distribution of T is not complete, therefore, many different estimators of $\bar{Y}$ can be suggested. For simplicity, we shall consider the following class of unbiased linear estimators of $\bar{Y}$.

$$\bar{y}_s = f_1(\nu)\, \bar{y}_\nu + f_2(\nu). \quad \ldots \quad (3.2)$$

In view of the fact that $\quad E[\bar{y}_s \mid \nu] = f_1(\nu)\bar{Y} + f_2(\nu)$,

obviously, necessary and sufficient conditions for $\bar{y}_s$ to be an unbiased estimator of $\bar{Y}$, are

$$E[f_1(\nu)] = 1 \quad \text{and} \quad E[f_2(\nu)] = 0. \quad \ldots \quad (3.3)$$

Consider now the class Γ of estimators $\bar{y}_s$ which satisfy the conditions of (3.3).

Now $$V(\bar{y}_s) = E\left[f_1^2(\nu)\left(\frac{1}{\nu} - \frac{1}{N}\right) S^2\right] + V[f_1(\nu)\bar{Y} + f_2(\nu)]. \quad \ldots \quad (3.4)$$

In order to choose a good estimator from Γ, we are to minimise (3.4) by proper choices of $f_1(\nu)$ and $f_2(\nu)$. The first expression on the right hand side of (3.4) is independent of $f_2(\nu)$; so, for a proper choice of $f_2(\nu)$, we are to minimise

$$V[f_1(\nu)\bar{Y} + f_2(\nu)]$$

which is minimum if $\bar{Y} f_1(\nu) + f_2(\nu)$ is constant for all values of ν, i.e.,

$$\bar{Y} f_1(\nu) + f_2(\nu) = E[f_1(\nu)\bar{Y} + f_2(\nu)] = \bar{Y},$$

or $$f_2(\nu) = \bar{Y}[1 - f_1(\nu)]. \quad \ldots \quad (3.5)$$

Since the above solution of $f_2(\nu)$ contains the unknown $\bar{Y}$, the exact value of $f_2(\nu)$ is not known unless $f_1(\nu) = 1$. Thus, if we choose $f_1(\nu) = 1$, the best estimator of $\bar{Y}$ would be $\bar{y}_\nu$. However, in practical situations, when some *a priori* knowledge about $\bar{Y}$ is available, it seems appropriate to approximate $f_2(\nu)$ by

$$f_2(\nu) = \bar{X}[1-f_1(\nu)], \qquad \ldots \quad (3.6)$$

where $\bar{X}$ is some *a priori* estimate of $\bar{Y}$. For example, $\bar{X}$ may be taken as the estimate of the population mean of the same variate obtained from some previous survey etc. On the other hand, if no such information about $\bar{Y}$ is available, it would be safe to take $f_2(\nu) = 0$. To choose the optimum value of $f_1(\nu)$, we have to minimise

$$Ef_1^2(\nu)\left(\frac{1}{\nu}-\frac{1}{N}\right)$$

subject to the condition that $E[f_1(\nu)] = 1$.

By Schwartz inequality we have

$$E\left[f_1^2(\nu)\left(\frac{1}{\nu}-\frac{1}{N}\right)\right].\, E\left[\left(\frac{1}{\nu}-\frac{1}{N}\right)^{-1}\right] \geqslant 1. \qquad \ldots \quad (3.7)$$

The equality holds if and only if

$$f_1(\nu) = \left(\frac{1}{\nu}-\frac{1}{N}\right)^{-1} \Big/ \left[E\left(\frac{1}{\nu}-\frac{1}{N}\right)^{-1}\right] = [N\nu/(N-\nu)]/E[N\nu/(N-\nu)]. \quad \ldots \quad (3.8)$$

Thus, when some *a priori* estimate $\bar{X}$ of $\bar{Y}$ is available, the optimum estimate of $\bar{Y}$ is given by

$$\bar{y}_{\nu(1)} = \frac{[N\nu/(N-\nu)]}{E[N\nu/(N-\nu)]}\,\bar{y}_\nu + \bar{X}\left[1-\frac{[N\nu/(N-\nu)]}{E[N\nu/(N-\nu)]}\right]. \qquad \ldots \quad (3.9)$$

When no such information about $\bar{Y}$ is available we may use the following estimator

$$\bar{y}_{\nu(2)} = \frac{[N\nu/(N-\nu)]}{E[N\nu/(N-\nu)]}\,\bar{y}_\nu. \qquad \ldots \quad (3.10)$$

The two estimators are admissible in Γ in the sense that they minimise the first component of (3.4). Any estimator $\bar{y}_s$ different from either of them cannot be uniformly better than $\bar{y}_{\nu(1)}$ or $\bar{y}_{\nu(2)}$ because

$V(\bar{y}_{\nu(1)}) < V(\bar{y}_s)$ for all populations where $\bar{Y} = \bar{X}$;

$V(\bar{y}_{\nu(2)}) < V(\bar{y}_s)$ for all populations where $\bar{Y} = 0$.

Expression for $E[N\nu/(N-\nu]$: Proceeding on similar lines given by the author (Pathak, 1961), it can be shown that

$$E[N\nu/(N-\nu)] = N^2 \sum_{m=1}^{n} \frac{p_{12\cdots m}}{(N-m)} \qquad \ldots \quad (3.11)$$

where $$p_{12\ldots m} = \begin{cases} 1-\binom{m}{1}(1-1/N)^n+\ldots(-)^m\binom{m}{m}(1-m/N)^n \text{ for } m \leqslant n; \\ 0 \quad \text{otherwise.} \end{cases}$$

Thus, we see that it may be quite cumbersome to compute the estimators (3.9) and (3.10) in case of large samples owing to the difficulty of computing $E\,[N\nu/(N-\nu)]$. If, however, the sampling fraction n/N can be ignored, the estimators reduce to

$$\bar{y}^*_{\nu(1)} = \frac{\nu}{E(\nu)}\,\bar{y}_\nu+\bar{X}\left[1-\frac{\nu}{E(\nu)}\right]; \qquad \ldots \quad (3.12)$$

$$\bar{y}^*_{\nu(2)} = \frac{\nu}{E(\nu)}\,\bar{y}_\nu. \qquad \ldots \quad (3.13)$$

It is easy to see that (3.13) is the well-known Horvitz-Thompson (1952) estimator in case of equal probability sampling. An interesting comparison between $V(\bar{y}^*_{\nu(2)})$ and $V(\bar{y}_\nu)$ is made below.

4. Comparison between $(V\bar{y}_\nu)$ and $V(\bar{y}^*_{\nu(2)})$

We have shown that

$$V(\bar{y}_\nu) = \frac{1^{n-1}+2^{n-1}+\ldots+(N-1)^{n-1}}{N^n}\,S^2;$$

and $$V(\bar{y}^*_{\nu(2)}) = E\left[V\left\{\frac{\nu}{E(\nu)}\bar{y}_\nu|\nu\right\}\right]+V\left[E\left\{\frac{\nu}{E(\nu)}\bar{y}_\nu|\nu\right\}\right]$$

$$= E\left[\frac{\nu^2}{E^2(\nu)}\left(\frac{1}{\nu}-\frac{1}{N}\right)S^2\right]+\frac{\bar{Y}^2}{E^2(\nu)}\,V(\nu)\,. \qquad \ldots \quad (4.1)$$

It can be shown that

$$E(\nu) = N\left[1-\left(1-\frac{1}{N}\right)^n\right];$$

$$E(\nu^2) = N\left[1-\left(1-\frac{1}{N}\right)^n\right]+N(N-1)\left[1-2\left(1-\frac{1}{N}\right)^n+\left(1-\frac{2}{N}\right)^n\right];$$

and $$V(\nu) = N\left(1-\frac{1}{N}\right)^n-N^2\left(1-\frac{1}{N}\right)^{2n}+N(N-1)\left(1-\frac{2}{N}\right)^n$$

$$\therefore \quad V(\bar{y}^*_{\nu(2)}) = \frac{S^2}{N^2\left[1-\left(1-\frac{1}{N}\right)^n\right]^2}\left[N\left\{1-\left(1-\frac{1}{N}\right)^n\right\}-\left\{1-\left(1-\frac{1}{N}\right)^n\right\}\right.$$

$$\left.-(N-1)\left\{1-2\left(1-\frac{1}{N}\right)^n+\left(1-\frac{2}{N}\right)^n\right\}\right]+\frac{\bar{Y}^2}{N^2\left[1-\left(1-\frac{1}{N}\right)^n\right]^2}$$

$$\times\left[N\left(1-\frac{1}{N}\right)^n-N^2\left(1-\frac{1}{N}\right)^{2n}+N(N-1)\left(1-\frac{2}{N}\right)^n\right] \ldots \quad (4.2)$$

Now

$$V(\bar{y}_\nu)-V(\bar{y}^*_{\nu(2)}) = S^2\left[\frac{1^{n-1}+2^{n-1}+\ldots+(N-1)^{n-1}}{N^n}-\frac{(N-1)}{N^2}\frac{\left[\left(1-\frac{1}{N}\right)^n-\left(1-\frac{2}{N}\right)^n\right]}{\left[1-\left(1-\frac{1}{N}\right)^n\right]^2}\right]$$

$$-\frac{\bar{Y}^2}{N}\cdot\frac{\left[\left(1-\frac{1}{N}\right)^n-N\left(1-\frac{1}{N}\right)^{2n}+(N-1)\left(1-\frac{2}{N}\right)^n\right]}{\left[1-\left(1-\frac{1}{N}\right)^n\right]^2}$$

$$= C_1S^2-C_2\bar{Y}^2. \quad \text{(say)} \qquad \ldots \quad (4.3)$$

Thus $\bar{y}_\nu$ is better than $\bar{y}^*_{\nu(2)}$ if

$$\frac{S^2}{\bar{Y}^2} < \frac{C_2}{C_1}$$

and worse if

$$\frac{S^2}{\bar{Y}^2} > \frac{C_2}{C_1}.$$

Approximate values of C_1 and C_2 for large populations correct up to terms of order N^{-2}, are given by

$$C_1 = \frac{1}{2nN}+\frac{5(n-1)}{12nN^2};$$

$$C_2 = \frac{(n-1)}{2nN}-\frac{(n-1)(n-2)}{3nN^2}. \qquad \ldots \quad (4.4)$$

The above comparison shows that if the square of the population coefficient of variation exceeds $(n-1)$, then $\bar{y}^*_{\nu(2)}$ has smaller variance than $\bar{y}_\nu$. Moreover, if we have some *a priori* knowledge of $\bar{Y}$, it would be more pertinent to compare $\bar{y}_\nu$ and $\bar{y}^*_{\nu(1)}$. It can be seen on similar lines that $\bar{y}_\nu$ is better than $\bar{y}^*_{\nu(2)}$ if

$$\frac{S^2}{(\bar{Y}-\bar{X})^2} < \frac{C_2}{C_1},$$

and worse otherwise. This result shows that if $\bar{X}$ provides a close approximation to $\bar{Y}$, it is always better to use $\bar{y}^*_{\nu(1)}$ rather than $\bar{y}_\nu$.

We now state the following admissibility property of $\bar{y}_\nu$.

Theorem 1 : *If squared error be the loss function, $\bar{y}_\nu$ is admissible among all functions of $\bar{y}_\nu$ and ν.*

Proof : Let $$t = \bar{y}_\nu+f(\bar{y}_\nu, \nu)$$
be a function of $\bar{y}_\nu$ and ν. Suppose that t is uniformly better than $\bar{y}_\nu$. Now by hypothesis,

$$R(t) = E(\bar{y}_\nu-\bar{Y})^2+E[f(\bar{y}_\nu, \nu)]^2+2E[(\bar{y}_\nu-\bar{Y})f(\bar{y}_\nu, \nu)] \leqslant E(\bar{y}_\nu-\bar{Y})^2 \quad \ldots \quad (4.5)$$

holds for all $Y_1, Y_2, \ldots, Y_N$. Take in particular $Y_1 = Y_2 = \ldots = Y_N = C$ (say). Then the above relation implies that

$$f(C, \nu) = 0. \qquad \ldots \quad (4.6)$$

Since the choice of C is arbitrary, it follows that $f(\bar{y}_\nu, \nu)$ is identically zero, which proves the above theorem.

5. Estimation of variance

We now turn to the problem of estimating the population variance from a simple random sample (with replacement). The usual estimator of the population variance

$$\sigma^2 = N^{-1} \Sigma (Y_j - \bar{Y})^2$$

is given by the sample variance

$$s^2 = \frac{1}{(n-1)} \Sigma (y_i - \bar{y})^2 = \frac{1}{2n(n-1)} \sum_{i \neq i'} (y_i - y_{i'})^2. \quad \dots \quad (5.1)$$

In this section, we derive an estimator uniformly better than s^2.

Theorem 2 : *For any convex (downwards) loss function, an estimator uniformly better than s^2 is given by*

$$s_\nu^2 = \left[\frac{C_\nu(n) - C_\nu(n-1)}{C_\nu(n)} \right] s_2^2, \quad \dots \quad (5.2)$$

where

$$C_\nu(n) = \nu^n - \binom{\nu}{1}(\nu-1)^n + \cdots (-)^{\nu-1} \binom{\nu}{\nu-1} 1^n,$$

and

$$s_2^2 = \begin{cases} \dfrac{1}{(\nu-1)} \Sigma (y_{(i)} - \bar{y}_\nu)^2 & \text{if } \nu > 1 ; \\ 0 & \text{otherwise.} \end{cases}$$

Proof : Since the 'order-statistic', T, is sufficient, by Rao-Blackwell theorem, an estimator uniformly better than s^2 is given by

$$E[s^2 | T] = E\left[\frac{1}{2n(n-1)} \sum_{i \neq i'} (y_i - y_{i'})^2 | T \right] = E[\tfrac{1}{2}(y_1 - y_2)^2 | T]. \quad \dots \quad (5.3)$$

When $\nu = 1$, (5.3) is obviously zero. To derive (5.3) when $\nu > 1$, we observe that

$$P[x_1 = x_{(i)}, x_2 = x_{(i')} | T] = \frac{\dfrac{1}{N^2} \Sigma''' \dfrac{(n-2)!}{\alpha_{(1)}! \dots \alpha_{(\nu)}!} \left(\dfrac{1}{N}\right)^{\alpha_{(1)}} \cdots \left(\dfrac{1}{N}\right)^{\alpha_{(\nu)}}}{\Sigma' \dfrac{n!}{\alpha_{(1)}! \dots \alpha_{(\nu)}!} \left(\dfrac{1}{N}\right)^{\alpha_{(1)}} \cdots \left(\dfrac{1}{N}\right)^{\alpha_{(\nu)}}} \dots \quad (5.4)$$

$$(i \neq i' = 1, 2, \dots, \nu)$$

where Σ' means summation over all integral α's such that

$$\alpha_{(1)} + \alpha_{(2)} + \cdots + \alpha_{(\nu)} = n \quad \text{and} \quad \alpha_{(i)} > 0 \quad \text{for} \quad i = 1, 2, \dots, \nu;$$

and Σ'' means summation over all integral α's such that

$\alpha_{(1)} + \alpha_{(2)} + \dots + \alpha_{(\nu)} = n-2$, $\alpha_{(i)} \geqslant 0$, $\alpha'_{(i)} \geqslant 0$ and $\alpha_{(k)} > 0$; for $k \neq i \neq i' = 1, 2, \dots, \nu$.

It follows from Lemma 1 given by the author (Pathak, 1961) that

$$\Sigma' \frac{n\,!}{\alpha_{(1)}\,!\ldots\alpha_{(\nu)}\,!} = C_\nu(n)\,;$$

$$\Sigma'' \frac{(n-2)!}{\alpha_{(1)}\,!\ldots\alpha_{(\nu)}\,!} = C_\nu(n-2)+2C_{\nu-1}(n-2)+C_{\nu-2}(n-2)$$

$$= \frac{C_\nu(n)-C_\nu(n-1)}{\nu(\nu-1)} \qquad \ldots \quad (5.5)$$

$$\therefore\ P[x_1 = x_{(i)},\quad x_2 = x_{(i')}\,|\,T] = \frac{C_\nu(n)-C_\nu(n-1)}{\nu(\nu-1)\ C_\nu(n)} \qquad \ldots \quad (5.6)$$

$$(i \neq i' = 1, 2, \ldots, \nu).$$

Thus, if $\nu > 1$

$$E\left[\frac{(y_1-y_2)^2}{2}\,|T\right] = \Sigma\,\frac{(y_{(i)}-y_{(i')})^2}{2}\,P[x_1 = x_{(i)},\, x_2 = x_{(i')}\,|\,T]$$

$$= \frac{C_\nu(n)-C_\nu(n-1)}{C_\nu(n)}\left[\frac{1}{2\nu(\nu-1)}\ \Sigma\,(y_{(i)}-y_{(i')})^2\right]$$

$$= \frac{C_\nu(n)-C_\nu(n-1)}{C_\nu(n)}\cdot\frac{1}{(\nu-1)}\ \Sigma\,(y_{(i)}-\bar{y}_\nu)^2. \qquad \ldots \quad (5.7)$$

Therefore, for any ν, $$E(s^2|T) = E\left[\frac{(y_1-y_2)^2}{2}\,|T\right] = \frac{C_\nu(n)-C_\nu(n-1)}{C_\nu(n)}\,s_d^2, \qquad \ldots \quad (5.8)$$

where s_d^2 has been defined earlier.

In practice the estimator s_ν^2 requires the knowledge of the ratio $\frac{C_m(n-1)}{C_m(n)}$ Table 3 gives values of $\frac{C_m(n-1)}{C_m(n)}$ correct to seven places of decimals for $1 \leqslant m \leqslant n \leqslant 50$; and were computed from values of $\frac{C_m(n)}{m\,!}$ tabulated by Gupta (1950).

The following results are direct consequences of Theorem 2.

If there are two characters Y and Z, the covariance between Y and Z is defined by

$$\sigma_{(yz)} = \frac{1}{N}\,\Sigma(Y_j-\bar{Y})(Z_j-\bar{Z}). \qquad \ldots \quad (5.9)$$

The usual estimator of σ_{yz} is given by

$$s_{(yz)} = \frac{1}{(n-1)}\ \Sigma(y_i-\bar{y})(z_i-\bar{z}). \qquad \ldots \quad (5.10)$$

Corollary 1 : *It follows from Theorem 2 that an estimator better than $s_{(yz)}$ is given by*

$$s_{r(yz)} = \begin{cases} \dfrac{C_\nu(n)-C_\nu(n-1)}{C_\nu(n)}\, s_{d(yz)} & \text{if } \nu > 1 \\ 0 & \text{otherwise,} \end{cases} \quad \dots \quad (5.11)$$

where $s_{d(yz)}$ is the sample covariance based on the distinct units observed in the sample.

The above theorem can be used to derive an unbiased ratio estimator which is better than Hartley-Ross unbiased ratio estimator (1954). In the sampling scheme under consideration, Hartley-Ross estimator is given by

$$\bar{y}_R = \bar{r}\bar{Z} - \frac{n}{(n-1)}(\bar{y}-\bar{r}\bar{z}) = \bar{r}\bar{Z} - \frac{1}{(n-1)} \Sigma \left(\frac{y_i}{z_i} - \bar{r} \right)(z_i - \bar{z}),$$

where $\bar{Z} = N^{-1} \Sigma Z_j$, $\bar{r} = 1/n\, \Sigma y_i/z_i$ and z_i is the value of the Z-characteristic, an auxiliary characteristic related to Y-characteristic, of the i-th sample unit.

Corollary 2 : *An estimator better than $\bar{y}_R$ is given by*

$$E[\bar{y}_R | T] = \begin{cases} \bar{r}_\nu \bar{Z} + \dfrac{C_\nu(n)-C_\nu(n-1)}{C_\nu(n)} \cdot \dfrac{\nu}{(\nu-1)} (\bar{y}_\nu - \bar{r}_\nu \bar{z}_\nu) & \text{if } \nu > 1 \\ \bar{r}_\nu \bar{Z} & \text{otherwise,} \end{cases} \quad \dots \quad (5.12)$$

where $\quad \bar{r}_\nu = \dfrac{1}{\nu} \Sigma \dfrac{y_{(i)}}{z_{(i)}}, \; \bar{z}_\nu = \dfrac{1}{\nu} \Sigma z_{(i)}.$

Murthy (1961) has extended the idea of ratio estimators to product estimators. Similar to the well-known ratio estimator $\frac{\bar{y}\bar{Z}}{\bar{z}}$ he has considered the product estimator

$$\frac{\bar{y}\bar{z}}{\bar{Z}} \quad \dots \quad (5.13)$$

for estimating $\bar{Y}$.

Corollary 3 : *It can be verified that an estimator better than $\frac{\bar{y}\bar{z}}{\bar{Z}}$ is given by*

$$E\left[\frac{\bar{y}\bar{z}}{\bar{Z}} \,\middle|\, T\right] = \frac{1}{\bar{Z}} \left[\frac{\Sigma\, y_{(i)} z_{(i)}}{\nu} - \frac{(n-1)}{n} \cdot s_{r(yz)} \right] \quad \dots \quad (5.14)$$

where $s_{r(yz)}$ is given by (5.11).

Finally, the almost unbiased product estimator of Murthy (1961), is given which makes $\frac{\bar{y}\bar{z}}{\bar{Z}}$ almost unbiased. This estimator is given by

$$P_e = \frac{n}{(n-1)} \cdot \frac{\bar{y}\bar{z}}{\bar{Z}} - \frac{1}{(n-1)} \frac{\Sigma y_i z_i}{n\bar{Z}}. \quad \dots \quad (5.15)$$

Corollary 4 : *An estimator better than P_e is given by*

$$E[P_e \mid T] = \frac{\Sigma x_{(i)}\, y_{(i)}}{\nu \bar{Z}} - \frac{s_{r(yz)}}{\bar{Z}}. \quad \dots \quad (5.16)$$

6. Some estimators of $V(\bar{y}_\nu)$

Some unbiased estimators of $V(\bar{y}_\nu)$ are given by

(I) $$v_1(\bar{y}_\nu) = \left[\frac{1^{n-1}+2^{n-1} \ldots +(N-1)^{n-1}}{N^n} \right] \frac{N}{(N-1)} s^2;$$

(II) $$v_2(\bar{y}_\nu) = \left[\frac{1^{n-1}+2^{n-1}+\ldots+(N-1)^{n-1}}{N^n} \right] \frac{N}{(N-1)} \frac{[C_\nu(n)-C_\nu(n-1)]}{C_\nu(n)} s_d^2 ;$$

(III) $$v_3(\bar{y}_\nu) = \frac{C_{\nu-1}(n-1)}{C_\nu(n)} s_d^2 ;$$

(IV) $$v_4(\bar{y}_\nu) = \left[\left(\frac{1}{\nu} - \frac{1}{N} \right) + \frac{(N-1)}{(N^n-N)} \right] s_d^2 ;$$

(V) $$v_5(\bar{y}_\nu) = \left[\left(\frac{1}{\nu} - \frac{1}{N} \right) + N^{1-n} \left(1 - \frac{1}{\nu} \right) \right] s_d^2 \quad \text{(to be used for } \nu > 1).$$

The estimate (II) is known to be uniformly better than (I). It appears difficult to give direct proofs of relative efficiencies of these estimators. The estimators (IV) and (V) were given by Des Raj and Khamis (1958). The estimator (V) is conditionally unbiased for $\nu > 1$. Des Raj and Khamis suggested the use of (V) for $\nu > 1$.

It is easy to see that

$$v_4 = v_5 \frac{N^n}{(N^n-N)}. \qquad \ldots \quad (6.1)$$

A little comparison will, now, show that the conditional variance of (V) is less than the variance of (IV). The amount of decrease in the variance is given by

$$V(v_4) - V(v_5/\nu > 1) = \frac{1}{N^{n-1}} E(v_4^2). \qquad \ldots \quad (6.2)$$

In general, this leads to the conclusion that any estimator $\hat{\sigma}^2$ of σ^2 which is unbiased for σ^2 and is equal to zero for $\nu = 1$, can be reduced to give a conditionally unbiased estimate of σ^2 for $\nu > 1$ whose conditional variance will be less than the variance of $\hat{\sigma}^2$. This conditionally improved estimator is related with $\hat{\sigma}^2$ by the following equation.

$$\hat{\sigma}^2 = \hat{\sigma}^2_{im} \left[\frac{N^n}{(N^n-N)} \right] \qquad \ldots \quad (6.3)$$

where σ^2_{im} stands for the conditionally improved estimator of σ^2.

Numerical example. To study the relative efficiency of the estimators of $V(\bar{y}_r)$, we consider the following three populations given by Yates and Grundy (1953).

TABLE 1. THREE POPULATIONS GIVEN BY YATES AND GRUNDY

population	A	B	C
unit	Y_j	Y_j	Y_j
1	0.5	0.8	0.2
2	1.2	1.4	0.6
3	2.1	1.8	0.9
4	3.2	2.0	0.8
ΣY_j	7.0	6.0	2.5

These populations were deliberately chosen by them as being more extreme than will be normally encountered in practice.

The table below gives variances of unbiased estimators of $V(\bar{y}_r)$ when $n = 3$. $V(v_1)$ is not given as $V(v_1) > V(v_2)$.

TABLE 2. VARIANCES OF UNBIASED ESTIMATORS OF $V(Y)$

population	$V(Y_r)$	$V(v_2)$	$V(v_3)$	$V(v_4)$	$V(v_5 \mid \nu > 1)$
A	0.29823	0.04940	0.05222	0.09017	0.07897
B	0.06125	0.00220	0.00232	0.00396	0.00348
C	0.020964	0.000279	0.000293	0.000490	0.000432

The results show that for the three populations

$$V(v_2) < V(v_3) < V(v_5 \mid \nu > 1) < V(v_4). \qquad \ldots \quad (6.4)$$

Thus v_2 appears to be most efficient estimator of $V(\bar{y}_r)$.

For $n = 2$, v_2 and v_3 are identical. The comparison thus strongly suggests the use of v_2 for estimating $V(\bar{y}_r)$.

For getting estimators of $V(t)$, where t is any unbiased estimator of $\bar{Y}$, the following procedure may be adopted.

$$v(t) = t^2 - \text{est }(\bar{Y}^2), \qquad \ldots \quad (6.5)$$

where est $(\bar{Y}^2)$ stands for an unbiased estimator of $\bar{Y}^2$ and can be obtained from any of the relations

$$\text{est }(\bar{Y}^2) = v_i(\bar{y}_r) - \bar{y}_r^2 \qquad (i = 1, 2, 3, 4, 5). \qquad \ldots \quad (6.6)$$

From the example considered, it is expected that

$$\text{est }(\bar{Y}^2) = v_i(\bar{y}_r) - \bar{y}_r^2 \qquad (i = 2, 3) \qquad \ldots \quad (6.7)$$

would fare better than the remaining estimators of $\bar{Y}^2$.

TABLE 3. VALUES OF $\frac{C_m(n-1)}{C_m(n)}$

$n \rightarrow$ m	2	3	4	5	6	7	8	9
1				1.000000 for all n				
2	0	.3333333	.4285714	.4666667	.4838710	.4920635	.4960630	.4980392
3		0	.1666667	.2400000	.2777778	.2990033	.3115942	.3193388
4			0	.1000000	.1538462	.1857143	.2057613	.2189189
5				0	.0666667	.1071429	.1333333	.1510574
6					0	.0476190	.0789474	.1005291
7						0	.0357143	.0606061
8							0	.0277778
9								0

$n \rightarrow$ m	10	11	12	13	14	15	16	17
1				1.000000 for all n				
2	.4990215	.4995112	.4997557	.4998779	.4999390	.4999695	.4999847	.4999924
3	.3242220	.3273569	.3293923	.3307253	.3316032	.3321838	.3325687	.3328243
4	.2278258	.2339966	.2383479	.2414585	.2437059	.2453432	.2465438	.2474286
5	.1634568	.1723544	.1788676	.1837118	.1873611	.1901391	.1922722	.1939216
6	.1159164	.1271791	.1355998	.1420028	.1469395	.1507901	.1538225	.1562302
7	.0785714	.0918937	.1010882	.1097724	.1158627	.1208858	.1245448	.1276595
8	.0480000	.0631313	.0747043	.0837155	.0908370	.0965357	.1011446	.1049062
9	.0222222	.0389610	.0518519	.0619607	.0700084	.0764970	.0817858	.0861370
10	0	.0181818	.0322581	.0433566	.0522416	.0594466	.0653539	.0702435
11		0	.0151515	.0271493	.0367965	.0446549	.0511277	.0565107
12			0	.0128205	.0231660	.0316239	.0386164	.0444536
13				0	.0109890	.0200000	.0274725	.0337299
14					0	.0095238	.0174419	.0240896
15						0	.0083333	.0153453
16							0	.0073529
17								0

TABLE 3. VALUES OF $\frac{C_m(n-1)}{C_m(n)}$ (Contd.)

$n \rightarrow$ / m	18	19	20	21	22	23	24	25
1				1.000000 for all n				
2	.4999962	.4999981	.4999990	.4999995	.4999998	.4999999	.4999999	.5000000
3	.3329943	.3331075	.3331828	.3332330	.3332665	.3332888	.3333036	.3333135
4	.2480833	.2485693	.2489308	.2492003	.2494015	.2495518	.2496643	.2497484
5	.1952045	.1962073	.1969942	.1976139	.1981031	.1984904	.1987974	.1990412
6	.1581551	.1597031	.1609542	.1619699	.1627974	.1634737	.1640281	.1644835
7	.1301924	.1322655	.1339720	.1353836	.1365502	.1375340	.1383520	.1390384
8	.1080005	.1105635	.1126991	.1144881	.1159936	.1172659	.1183450	.1192632
9	.0897460	.0927608	.0952950	.0974369	.0992561	.1008081	.1021373	.1032797
10	.0743243	.0777549	.0806574	.0831272	.0852394	.0870540	.0886193	.0899745
11	.0610250	.0648389	.0680820	.0708560	.0732409	.0753009	.0770878	.0786436
12	.0493679	.0535361	.0570949	.0601513	.0627902	.0650794	.0670738	.0688180
13	.0390147	.0435116	.0473637	.0506832	.0535590	.0560624	.0582511	.0601724
14	.0297189	.0345220	.0386476	.0422127	.0453100	.0480141	.0503852	.0524727
15	.0212963	.0263854	.0307669	.0345620	.0378671	.0407595	.0433020	.0455460
16	.0136054	.0189618	.0235845	.0275957	.0310960	.0341657	.0368697	.0392613
17	.0065359	.0121457	.0169935	.0212082	.0248926	.0281295	.0309861	.0335172
18	0	.0058480	.0109091	.0153161	.0191746	.0225696	.0255705	.0282338
19		0	.0052632	.0098522	.0138756	.0174206	.0205584	.0233471
20			0	.0047619	.0089419	.0126294	.0158973	.0188053
21				0	.0043290	.0081522	.0115440	.0145657
22					0	.0039526	.0074627	.0105029
23						0	.0036232	.0068571
24							0	.0033333
25								0

$n \rightarrow$ / m	26	27	28	29	30	31	32	33
1				1.0000000 for all n				
2				.5000000 for $n > 24$				
3	.3333201	.3333245	.3333275	.3333294	.3333307	.3333316	.3333322	.3333326
4	.2498115	.2498587	.2498940	.2499206	.2499404	.2499553	.2499665	.2499749
5	.1992352	.1993695	.1995125	.1996106	.1996889	.1997514	.1998012	.1998411
6	.1648585	.1651676	.1654230	.1656342	.1658090	.1659539	.1660741	.1661738
7	.1396157	.1401025	.1405137	.1408616	.1411565	.1414068	.1416195	.1418004
8	.1200468	.1207173	.1212923	.1217864	.1222119	.1225788	.1228958	.1231699
9	.1042646	.1051161	.1058542	.1064955	.1070539	.1075409	.1079665	.1083390
10	.0911518	.0921776	.0930737	.0938586	.0945476	.0951537	.0956879	.0961594
11	.0800030	.0811943	.0822415	.0831643	.0839795	.0847012	.0853413	.0859103
12	.0703490	.0716970	.0728875	.0739417	.0748775	.0757102	.0764525	.0771157
13	.0618648	.0633607	.0646867	.0658656	.0669162	.0678548	.0686950	.0694488
14	.0543172	.0559625	.0574068	.0587037	.0598633	.0609026	.0618362	.0626766
15	.0475339	.0493009	.0508763	.0522851	.0535481	.0546833	.0557058	.0566280
16	.0413846	.0432760	.0449662	.0464810	.0478422	.0490684	.0501756	.0511775
17	.0357685	.0377778	.0395768	.0411922	.0426467	.0439595	.0451473	.0462243
18	.0306065	.0327276	.0346298	.0363408	.0378840	.0392793	.0405439	.0416926
19	.0258351	.0280625	.0300629	.0318649	.0334925	.0349665	.0363044	.0375215
20	.0214030	.0237315	.0258255	.0277142	.0294224	.0309714	.0323703	.0336619
21	.0172679	.0196929	.0218762	.0238477	.0256329	.0272536	.0287284	.0300736
22	.0133950	.0159121	.0181807	.0202314	.0220902	.0237795	.0253185	.0267237
23	.0097547	.0123601	.0147103	.0168368	.0187662	.0205213	.0221217	.0235844
24	.0063224	.0090123	.0114409	.0136400	.0156371	.0174553	.0191148	.0206327
25	.0030769	.0058480	.0083516	.0106206	.0126827	.0145616	.0162777	.0178488
26	0	.0028490	.0054250	.0077611	.0098857	.0118229	.0135936	.0152159
27		0	.0026455	.0050463	.0072310	.0092245	.0110478	.0127193
28			0	.0024631	.0047059	.0067535	.0086276	.0103467
29				0	.0022989	.0043988	.0063218	.0080868
30					0	.0021505	.0041209	.0059303
31						0	.0020161	.0038685
32							0	.0018939
33								0

TABLE 3. VALUES OF $\frac{C_m(n-1)}{C_m(n)}$ (*Contd.*)

$n \rightarrow$ m	34	35	36	37	38	39	40	41	42
1				1.0000000	for	all n			
2				.5000000	for	$n>24$			
				.3333333	for	$n>38$			
3	.3333328	.3333330	.3333331	.3333332	.3333332	.3333333	.3333333	for $n>38$	
4	.2490812	.2499850	.2499894	.2499921	.2400940	.2499955	.2499966	.2490975	.2490981
5	.1998730	.1998984	.1999188	.1999350	.1999480	.1999584	.1999667	.1999734	.1999787
6	.1662566	.1663255	.1663827	.1664302	.1664698	.1665027	.1665301	.16615530	.1665720
7	.1419544	.1420856	.1421076	.1422031	.1423746	.1424442	.1425037	.1425546	.1425981
8	.1234073	.1236131	.1237916	.1239467	.1240815	.1241988	.1243008	.1243897	.1244671
9	.1086054	.1089519	.1092036	.1094250	.1096199	.1097916	.1099431	.1100767	.1101948
10	.0965764	.0969451	.0972731	.0975637	.0978221	.0980519	.0982566	.0984390	.0986017
11	.0864167	.0868683	.0872715	.0876321	.0879548	.0882441	.0885036	.0887367	.0889462
12	.0777093	.0782413	.0787190	.0791485	.0795352	.0798837	.0801982	.0804824	.0807393
13	.0701262	.0707361	.0712861	.0717828	.0722320	.0726388	.0730075	.0733423	.0736464
14	.0634346	.0641193	.0647390	.0653007	.0658105	.0662739	.0666957	.0670799	.0674305
15	.0574637	.0582202	.0589068	.0595311	.0600994	.0606176	.0610907	.0615232	.0619189
16	.0520858	.0529109	.0536617	.0543460	.0549707	.0555417	.0560643	.0565434	.0569830
17	.0472028	.0480935	.0489058	.0496476	.0503263	.0509481	.0515186	.0520426	.0525245
18	.0427382	.0436916	.0445626	.0453597	.0460903	.0467608	.0473773	.0479446	.0484674
19	.0386311	.0396445	.0405718	.0414218	.0422021	.0429196	.0435802	.0441893	.0447515
20	.0348327	.0359036	.0368848	.0377855	.0386136	.0393761	.0400792	.0407284	.0413286
21	.0313031	.0324290	.0334620	.0344113	.0352853	.0360910	.0368350	.0375228	.0381596
22	.0280093	.0291880	.0302707	.0312668	.0321848	.0330321	.0338154	.0345404	.0352124
23	.0249241	.0261535	.0272838	.0283248	.0292852	.0301725	.0309937	.0317545	.0324605
24	.0220242	.0233023	.0244785	.0255627	.0265639	.0274898	.0283474	.0291428	.0298815
25	.0192902	.0206152	.0218354	.0229613	.0240017	.0249648	.0258575	.0266863	.0274566
26	.0167052	.0180754	.0193382	.0205041	.0215825	.0225813	.0235080	.0243689	.0251697
27	.0142550	.0156687	.0169725	.0181771	.0192920	.0203255	.0212849	.0221769	.0230072
28	.0119270	.0133827	.0147261	.0159681	.0171183	.0181852	.0191763	.0200983	.0209572
29	.0097103	.0112066	.0125883	.0138664	.0150508	.0161500	.0171717	.0181228	.0190092
30	.0075954	.0091310	.0105497	.0118628	.0130802	.0142107	.0152620	.0162412	.0171545
31	.0055740	.0071475	.0086021	.0099490	.0111984	.0123592	.0134393	.0144458	.0153849
32	.0036386	.0052489	.0067382	.0081179	.0093983	.0105885	.0116964	.0127294	.0136937
33	.0017825	.0034286	.0049515	.0063630	.0076735	.0088922	.0100272	.0110858	.0120745
34	0	.0016807	.0032362	.0046786	.0060183	.0072647	.0084259	.0095094	.0105218
35		0	.0015873	.0030597	.0044277	.0057010	.0068877	.0079955	.0090309
36			0	.0015015	.0028972	.0041965	.0054081	.0065394	.0075972
37				0	.0014225	.0027473	.0039820	.0051372	.0062168
38					0	.0013495	.0026087	.0037853	.0048862
39						0	.0012821	.0024804	.0036020
40							0	.0012195	.0023613
41								0	.0011614
42									0

TABLE 3. VALUES OF $\frac{C_m(n-1)}{C_m(n)}$ (*Contd.*)

$n \rightarrow$ m	43	44	45	46	47	48	49	50
1				1.0000000 for all n				
2				.5000000 for $n > 24$				
3				.3333333 for $n > 38$				
4	.2499986	.2499989	.2499992	.2499994	.2499996	.2499997	.2499997	.2400998
5	.1999830	.1999864	.1990891	.1999913	.1999930	.1999944	.1999955	.1999964
6	.1665878	.1666009	.1666119	.1666210	.1666287	.1666350	.1666403	.1666447
7	.1426354	.1426672	.1426945	.1427178	.1427378	.1427540	.1427695	.1427821
8	.1245346	.1245935	.1246897	.1246897	.1247288	.1247629	.1247928	.1248188
9	.1102991	.1103913	.1104729	.1105451	.1106090	.1106656	.1107158	.1107602
10	.0987470	.0988767	.0989927	.0990965	.0991893	.0992724	.0993468	.0994135
11	.0891346	.0893043	.0894572	.0895950	.0897194	.0898316	.0899330	.0900246
12	.0809719	.0811826	.0813736	.0815469	.0817042	.0818471	.0810771	.0820953
13	.0739230	.0741749	.0744043	.0746136	.0748045	.0749789	.0751383	.0752840
14	.0677505	.0680431	.0683107	.0685558	.0687804	.0689864	.0691754	.0693400
15	.0622814	.0626139	.0629191	.0631995	.0634574	.0636947	.0639132	.0641147
16	.0573868	.0577581	.0581000	.0584150	.0587054	.0589736	.0592212	.0594502
17	.0529683	.0533774	.0537549	.0541035	.0544259	.0547241	.0550003	.0552563
18	.0489498	.0493954	.0498074	.0501888	.0505420	.0508696	.0511736	.0514560
19	.0452712	.0457520	.0461974	.0466104	.0469937	.0473498	.0476809	.0479890
20	.0418842	.0423991	.0428768	.0433204	.0437328	.0441165	.0444738	.0448069
21	.0387498	.0392976	.0398064	.0402796	.0407202	.0411307	.0415135	.0418708
22	.0358361	.0364155	.0369545	.0374563	.0379240	.0383604	.0387679	.0391488
23	.0331164	.0337264	.0342944	.0348239	.0353180	.0357794	.0362108	.0366145
24	.0305684	.0312080	.0318041	.0323603	.0328798	.0333656	.0338201	.0342459
25	.0281735	.0288416	.0294649 •	.0300469	.0305911	.0311003	.0315773	.0320245
26	.0259157	.0266113	.0272608	.0278679	.0284359	.0289679	.0294666	.0299345
27	.0237812	.0245035	.0251784	.0258097	.0264008	.0269548	.0274745	.0279626
28	.0217583	.0225064	.0232059	.0238606	.0244740	.0250494	.0255895	.0260971
29	.0198366	.0206097	.0213330	.0220104	.0226455	.0232416	.0238016	.0243281
30	.0180073	.0188046	.0195510	.0202504	.0209065	.0215227	.0221019	.0226468
31	.0162624	.0170832	.0178520	.0185728	.0192493	.0198850	.0204828	.0210456
32	.0145950	.0154386	.0162291	.0169707	.0176670	.0183216	.0189376	.0195177
33	.0129991	.0138648	.0146764	.0154381	.0161537	.0168267	.0174603	.0180573
34	.0114690	.0123563	.0131884	.0139690	.0147040	.0153949	.0160456	.0166590
35	.0100000	.0109082	.0117602	.0125605	.0133131	.0140214	.0146887	.0153181
36	.0085877	.0095162	.0103876	.0112065	.0119767	.0127020	.0133856	.0140305
37	.0072281	.0081762	.0090667	.0099036	.0106911	.0114329	.0121323	.0127924
38	.0059176	.0068852	.0077939	.0086484	.0094527	.0102106	.0109254	.0116003
39	.0046531	.0056395	.0065662	.0074378	.0082585	.0090321	.0097619	.0104512
40	.0034317	.0044364	.0053806	.0062689	.0071056	.0078944	.0086390	.0093423
41	.0022506	.0032731	.0042344	.0051391	.0059913	.0067951	.0075540	.0082710
42	.0011074	.0021475	.0031254	.0040460	.0049135	.0057319	.0065047	.0072351
43	0	.0010571	.0020513	.0029874	.0038698	.0047024	.0054889	.0062324
44		0	.0010101	.0019614	.0028584	.0037049	.0045047	.0052611
45			0	.0009662	.0018773	.0027375	.0035504	.0043192
46				0	.0009251	.0017986	.0026242	.0034053
47					0	.0008865	.0017246	.0025177
48						0	.0008503	.0016552
49							0	.0008163
50								0

• *Note* : Values of $\frac{C_{m-1}(n-1)}{C_m(n)}$ can also be obtained from $\frac{C_m(n-1)}{C_m(n)}$ by using following relation

$$\frac{1}{m} = \frac{C_m(n-1)}{C_m(n)} + \frac{C_{m-1}(n-1)}{C_m(n)}$$

7. Comparison between with and without replacement simple random sampling schemes

In conclusion let us compare the two simple random sampling schemes for the purpose of estimation of $\bar{Y}$. If we draw a simple random sample with replacement of size n, then the variance of the sample mean is σ^2/n. Further, in a simple random sample without replacement of size n, the variance of the sample mean is $\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$. Since

$$\frac{\sigma^2}{n}\cdot\frac{(N-n)}{(N-1)} < \frac{\sigma^2}{n}$$

it is usually claimed that sampling without replacement is better than sampling with replacement. Basu (1958) has pointed out that this comparison is not fair because the cost of selecting a sample of size n in sampling without replacement is greater than the cost of selecting a sample in sampling with replacement. For comparing the two sampling schemes, it would be appropriate to take into account the cost involved in the selection of two different samples. The comparison, thus, mainly depends on the choice of the cost function, and no sampling scheme can be said to be superior to the other unless the cost function is known in advance. Let us, for illustration, consider the case where the cost of sampling is proportional to the number of distinct units drawn. Thus the expected cost of selecting a sample with replacement of size n is equivalent to the cost of selecting a sample without replacement of size $E(\nu) = N\left[1-\left(\frac{N-1}{N}\right)^n\right]$. Basu has shown that in this situation the sample mean of the sample with replacement is worse than the sample mean of the equivalent sample without replacement. We now compare the sample mean $\bar{y}$ of the equivalent sample without replacement with the following estimator of with replacement sample :

$$\bar{y}_{\nu(2)} = \frac{[N\nu/(N-\nu)]}{E[N\nu/(N-\nu)]}\bar{y}_\nu.$$

It has been shown that

$$V(\bar{y}_{\nu(2)}) = \frac{S^2}{E[N\nu/(N-\nu)]} + \bar{Y}^2 V\left[\frac{N\nu/(N-\nu)}{E\{N\nu/(N-\nu)\}}\right], \qquad \dots \quad (7.1)$$

and

$$V(\bar{y}) = \left[\frac{1}{E(\nu)} - \frac{1}{N}\right]S^2. \qquad \dots \quad (7.2)$$

Since $N\nu/(N-\nu)$ is a convex function of $\nu (1 \leqslant \nu \leqslant n < N)$,

$$E\,[N\nu/(N-\nu)] \geqslant NE\nu/[N-E\nu] = \left[\frac{1}{E\nu} - \frac{1}{N}\right]^{-1} \qquad \dots \quad (7.3)$$

From (7.3), it is evident that the first component of $V(\bar{y}_{\nu(2)})$ is smaller than $V(\bar{y})$. Thus for a population whose coefficient of variation is sufficiently large $V(\bar{y}_{\nu(2)})$ would be smaller than $V(\bar{y})$. This comparison shows that the sample mean of without replacement sample cannot be uniformly better than all estimators of with replacement sampling.

However, the comparison made above is not very satisfactory. First, because of the linearity of the cost function and secondly, because $E(\nu)$ is not necessarily an integer. We hope that for some other cost functions also, similar situations may be found out where with replacement sampling would fare better than without replacement sampling.

ACKNOWLEDGEMENT

I am grateful to Dr. D. Basu for his valuable guidance in the preparation of the paper.

REFERENCES

BASU, D. (1958): On sampling with and without replacement. *Sankhyā*, **20**, 287-294.

DES RAJ and KHAMIS, S. H. (1958): Some remarks on sampling with replacement. *Ann. Math. Stat.*, **29**, 550-557.

FRASER, D. A. S. (1957): *Non-parametric Methods in Statistics*, John Wiley and Sons, New York.

GUPTA, H. (1950): Tables of distributions. *Research Bulletin of the East Punjab University*, No. 2, 3. 13-44.

HARTLEY, H. O. and ROSS, A. (1954): Unbiased ratio estimators. *Nature*, **174**, 270-271.

HORVITZ, D. G. and THOMPSON, D. J. (1952): A generalization of sampling without replacement from finite universes. *J. Amer. Stat. Ass.*, **47**, 663-685.

MURTHY, M. N. (1961): Some problems of estimation in sampling from finite populations. Thesis submitted for the Ph.D degree of the Indian Statistical Institute, 105-112.

PATHAK, P. K. (1961): On the evaluation of moments of distinct units in a sample, *Sankhyā*, Series A, **23**, 415-420

YATES, F. and GRUNDY, P. M. (1953): Selection without replacement from within strata with probability proportional to size. *J. Roy. Stat. Soc.*, Series B, **15**, 253-261.

Paper received : September, 1961.