Jour. Ind. Soc. Ag. Statistics
54(2), 2001 : 355–369

On the Estimation of Population Variance in Repeat Surveys

U.C. Sud, A.K. Srivastava and D.P. Sharma
Indian Agricultural Statistics Research Institute, New Delhi - 110012
(Received : March, 2001)

SUMMARY

A general theory of estimation of population variance in case of repeat surveys is presented. Estimator of population variance for the current occasion is obtained as a particular case for sampling on two occasions. Expressions for optimum proportion of matched (λ) and unmatched (μ) units are obtained. Substantial gain in precision is observed, when the optimum value of λ and μ are used, over an estimator that does not utilise previous occasion information. However, when the optimum proportion of matched (say, λ') and unmatched (μ') units for estimating population mean are used the loss in precision of an estimator of population variance is marginal. This suggests that, in case of repeat surveys, the problem of estimation of population variance can be simultaneously considered with the estimation of population mean.

Key words: Repeat survey, Population mean, Population variance.

1. Introduction

Jessen [3], Tikkiwal [4], Patterson [4], Eckler [2] and Raj [5] contributed towards the development of the theory of unbiased estimation of mean of characteristics in successive sampling. This approach is extended to develop general theory of estimation for repeat surveys for the population variance.

2. Some General Results

Consider a finite population $\Omega = (U_1, U_2, .., U_N)$ of 'N' identifiable units. Let 'y' be the variable under study taking value y_i on unit U_i, i = 1,2,.., N. Let the parameter of interest be $\sigma^2 = \frac{1}{N}\sum_{k}^{N} \left(y_k - \overline{Y}_N\right)^2$ where $\overline{Y}_N = \frac{1}{N}\sum_{k}^{N} y_k$

To estimate σ^2 a sample of size 'n' units is drawn. We assume that the sampling is from an infinite population. Consider estimator of the type

$$\hat{T}_Q = y'Ay = \sum_{k}^{n} \sum_{i}^{n} l_{kj} y_k y_j \tag{2.1}$$

where y is (n × 1) column vector with elements

$$y = (y_1, y_2 \ldots, y_n)'$$

and A is a symmetric semi-positive definite matrix of order (n × n) comprising of elements l_{kj}, k = 1, ..., n; j = 1, ..., n. It may be mentioned that the assumption that A is a symmetric matrix does not effect the generality of results, as any quadratic form $y'By$ can be written as $y'By = y'Ay$, where $A = \frac{1}{2}(B + B')$ is symmetric. It may be further noted that the assumption of semi positive definiteness of A matrix will ensure the non-negativity of the estimator which is a desirable property for estimation of variance.

Following result will be used in the proof of Theorems 1 and 2.

Lemma: If the sum of all elements of a symmetric semi-positive definite matrix is equal to '0' then individual row and column sum are equal to '0', (Vijayan, [6]).

We prove the following theorem.

Theorem 1. $\hat{T}_Q$ is a minimum variance unbiased estimator of σ^2 if and only if $\mathrm{Cov}\left(y_i^2, \hat{T}_Q\right) = \lambda_1$ and $\mathrm{Cov}\left(y_i y_j, \hat{T}_Q\right) = \lambda_2$, where λ_1 and λ_2 are the Lagrangian multipliers.

Proof: The condition is necessary

Unbiasedness of $\hat{T}_Q$ implies that

$$\sum_{k}^{n} l_{kk} = 1 \text{ and } \sum_{k}^{n} \sum_{j}^{n} l_{kj} = 0, \text{ or } \sum_{k} l_{kk} = 1 \text{ and } \sum_{k} \sum_{j(\neq k)}^{n} l_{kj} = -1$$

We minimize the variance of $\hat{T}_Q$ subject to the conditions of unbiasedness, in other words we minimize

$$\phi = V\left(\hat{T}_Q\right) - 2\lambda_i \left(\sum_{k} l_{kk} - 1 \right) - 2\lambda_2 \left(\sum_{k} \sum_{j(\neq k)} l_{kj} + 1 \right)$$

It can be seen that

$$V\left(\hat{T}_Q\right)=\mu_4\sum_{k} l_{kk}^2+2\left(\sigma^2\right)^2\sum_{k}\sum_{j} l_{kj}^2+4\overline{Y}_N\mu_3\sum_{k}\sum_{j} l_{kk}l_{kj}$$

$$+4\sigma'\overline{Y}_N'\sum_{k}\sum_{j}\sum_{j'} l_{kj}l_{kj'}+3\left(\sigma^2\right)'\sum_{k} l_{kk}' \qquad (2.2)$$

where $\mu_r=\frac{1}{N}\sum_{k}^{N}\left(y_k-\overline{Y}_N\right)^r$

Now, minimizing ϕ and using the lemma, we get the desired result as

$$\mu_4 l_{kk}+2\overline{Y}_N\mu_3 l_{kk}-\left(\sigma^2\right)^2 l_{kk}=\lambda_1=\text{Cov}\left(y_i^2,\hat{T}_Q\right) \qquad (2.3)$$

$$2\left(\sigma^2\right)^2 l_{kj}+\overline{Y}_N\mu_3 l_{kk}+\overline{Y}_N\mu_3 l_{jj}=\lambda_2=\text{Cov}\left(y_i y_j,\hat{T}_Q\right) \qquad (2.4)$$

The sufficiency condition can be proved as follows

Let $\hat{T}_Q$ be an unbiased estimator of σ^2 such that $\text{Cov}\left(y_i^2,\hat{T}_Q\right)=\lambda_1$ and $\text{Cov}\left(y_i y_j,\hat{T}_Q\right)=\lambda_2$, implying thereby

$$\text{Cov}\left(\sum_{k}\xi_k y_k^2,\hat{T}_Q\right)=\lambda_1 \text{ and } \text{Cov}\left(\sum_{k}\sum_{j\neq(k)}\xi_{kj}y_k y_j,\hat{T}_Q\right)=-\lambda_2$$

where $\sum_{k}\xi_k=1$ and $\sum_{k}\sum_{j(\neq k)}\xi_{kj}=-1$

Then, $\text{Cov}\left(\hat{T}_Q',\hat{T}_Q\right)=\lambda_1-\lambda_2$ (2.5)

where $\hat{T}_Q'=\sum_{k}\sum_{j}\xi_{kj}y_k y_j$ is another unbiased estimator of σ^2.

The equality (2.5) implies that

$$\text{Cov}\left(\hat{T}_Q'-\hat{T}_Q,\hat{T}_Q\right)=0, \text{ or } E\left(\hat{T}_Q',\hat{T}_Q\right)=E\left(\hat{T}_Q^2\right)$$

which further implies that

$$\mathrm{Cov}\left(\hat{T}'_Q, \hat{T}_Q\right) = V\left(\hat{T}_Q\right) \text{ or } \rho_{\hat{T}_Q \hat{T}'_Q} \sqrt{V\left(\hat{T}_Q\right) V\left(\hat{T}'_Q\right)} = V\left(\hat{T}_Q\right)$$

$$\sqrt{V\left(\hat{T}_Q\right)} = \rho_{\hat{T}_Q \hat{T}'_Q} \sqrt{V\left(\hat{T}'_Q\right)} \tag{2.6}$$

where $\rho_{\hat{T}_Q, \hat{T}'_Q}$ is the correlation coefficient between $\hat{T}_Q$ and $\hat{T}'_Q$.

From (2.6) we can see that $V\left(\hat{T}_Q\right) \le V\left(\hat{T}'_Q\right)$.

This completes the proof of the theorem.

Solving (2.3) and (2.4) for l_{kk} and l_{kj} we get

$$l_{kk} = \frac{1}{n}; \; l_{kj} = -\frac{1}{n(n-1)}$$

Substituting the optimum values of l_{kk} and l_{kj} in (2.1) and (2.2), we get

$$\hat{T}_{Qopt} = \frac{1}{(n-1)} \sum_k (y_k - \bar{y}_n)^2$$

and $$V\left(\hat{T}_{Qopt}\right) = \frac{1}{n}\left[\mu_4 - \frac{(n-3)}{(n-1)}\left(\sigma^2\right)^2\right] = \lambda_{1opt} - \lambda_{2opt}$$

where λ_{1opt} and λ_{2opt} are obtained by substituting the optimum values of l_{kk} and l_{kj} in (2.3) and (2.4) respectively.

2.1 Sampling on Successive Occasions with Partial Replacement of Units

We now assume that there are number of variates divided into different sets. The j^{th} variate in i^{th} set is denoted by $y_{ij}; (1 \le i \le h, 1 \le j \le n)$, where i is the time identifier and j the population element. We call it a sample pattern P. This P can be visualized as an incomplete matrix with m rows and the number of columns equal to the number of distinct elements in the sample. The population variance in the m^{th} set is denoted by

$$\sigma_m^2 = \frac{1}{N} \sum_i^N \left(y_{im} - \bar{Y}_m\right)^2$$

and its estimate by

$$\hat{T}_{QM} = \sum_{i}^{m} y_i' A_i y_i$$

where, y_i is $(n \times 1)$ column vector given by

$$y_i = (y_{i1}, y_{i2}, \ldots, y_{in})$$

and A_i is $(n \times n)$ symmetric semi-positive definite matrix comprising of elements ξ_{ikj}, $k = 1, \ldots, n; j = 1, \ldots, n$ such that

$$\sum_{k}^{n} \xi_{ikk} = 1; \quad \sum_{k}^{n} \sum_{j \neq (k)}^{n} \xi_{ikj} = -1, \ \forall i$$

We now give the following theorem

Theorem 2. $\hat{T}_{Qm}$ is a minimum variance unbiased estimator of σ_m^2 if and only if $\mathrm{Cov}\left(y_{im}^2, \hat{T}_{Qm}\right) = \lambda_{1m;}\ \mathrm{Cov}\left(y_{im}\, y_{jm}, \hat{T}_{Qm}\right) = \lambda_{2m}$, where λ_{1m} and λ_{2m} are the Lagrangian multipliers. The necessary condition can be proved as follows

Unbiasedness of $\hat{T}_{Qm}$ implies that

$$\sum_{k} \xi_{mkk} = 1, \sum_{k} \sum_{k(\neq j)} \xi_{mjk} = -1$$

and

$$\sum_{k} \xi_{ikk} = 0, \sum_{k} \sum_{k(\neq j)} \xi_{ikj} = 0, \forall\, i \neq m$$

By definition, the variance of $\hat{T}_{Qm}$ is

$$V\left(\hat{T}_{Qm}\right) = \sum_{i} V(y_i' A_i y_i) + \sum_{i \neq j} \mathrm{Cov}\left(y_i' A_i y_i, y_j' A_j y_j\right)$$

where the covariance terms pertain to units common to i^{th} and j^{th} occasion.

The expression for $V(y_i' A_i y_i)$ can be obtained on the similar lines as in Section 2. Also,

$$\mathrm{Cov}\left(y_i' A_i y_i, y_j' A_j y_j\right) = \mu_{2y_i,2y_j} \sum_k \xi_{ikk} \xi_{jkk} + \overline{Y}_i \mu_{y_i,2y_j} 2 \sum_k \sum_{k'} \xi_{ikk'} \xi_{jkk}$$

$$- \mu_{2y_i} \mu_{2y_j} \sum_k \xi_{ikk} \xi_{jkk} + 2 \overline{Y}_j \mu_{2y_i, y_j} \sum_k \sum_{k'} \xi_{ikk} \xi_{jkk'}$$

$$+ 2\mu^2_{y_i y_j} \sum_k \sum_{k'(\neq k)} \xi_{ikk'} \xi_{jkk'} + 4 \overline{Y}_i \overline{Y}_j \mu_{y_i y_j} \sum_k \sum_{k'} \sum_l \xi_{ikk'} \xi_{jkl}$$

where $\mu_{ry_i, sy_j} = \frac{1}{N} \sum_l \left(y_{il} - \overline{Y}_i\right)^r \left(y_{jl} - \overline{Y}_j\right)^s ; \overline{Y}_i = \frac{1}{N} \sum_j y_{ij}$

Minimizing $V\left(\hat{T}_{Qm}\right)$ subject to the condition of unbiasedness, i.e minimizing

$$V\left(\hat{T}_{Qm}\right) - 2 \sum_i \lambda_{1i} \left(\sum_k \xi_{ikk} - 1 \right) - 2 \sum_i \lambda_{2i} \left(\sum_k \sum_{j(\neq k)} \xi_{ikj} + 1 \right)$$

where λ's are the Lagrangian multipliers, gives the desired necessary condition.

Sufficiency can be proved on the similar lines as in Section 2.

The following corollary of Theorem 2 is useful.

Corollary. Let $\hat{T}'_{Qm}, \hat{T}_{Qm}$ be respectively unbiased and minimum variance unbiased estimator of σ^2_m. Then $\mathrm{Cov}\left(\hat{T}'_{Qm}, \hat{T}_{Qm}\right) = \lambda_{1m_{opt}} - \lambda_{2m_{opt}} = V_{min}\left(\hat{T}_{Qm}\right)$

It can be seen that many covariance conditions need to be checked up to determine whether or not $\hat{T}_{Qm}$ is minimum-variance. These can be reduced by the following theorem.

Theorem 3. Assume that the conditions of Theorem 2 hold. Let the sample pattern, denoted by P, be broken up into a finite number of rectangular sub-patterns: Consider the sub-pattern P_i which forms a complete matrix of y_{ij} values with r rows and c columns. Then the c weights ξ_{ikk}, ξ_{ikj} associated with the values of y^2_{ik} and $y_{ik}\, y_{ij}$ respectively in any one of the rows are equal i.e.

$$\xi_{111} = \xi_{122} = \ldots = \xi_{1cc}; \xi_{211} = \xi_{222} = \ldots = \xi_{2cc}; \ldots, \xi_{r11} = \xi_{r22} = \ldots = \xi_{rcc} \text{ and} \quad (2.6)$$

$$\xi_{112} = \xi_{113} = \ldots = \xi_{1c,c-1}; \xi_{212} = \xi_{213} = \ldots = \xi_{2c,c-1}; \ldots;$$

$$\xi_{r12} = \xi_{r13} = \ldots = \xi_{rc,c-1} \quad (2.7)$$

Proof: According to Theorem 2, the covariance condition must hold for this sub-pattern. Consider first the identity

$$\text{Cov}\left(y_{11}^2, \hat{T}_{Qm}\right) - \text{Cov}\left(y_{12}^2, \hat{T}_{Qm}\right) = 0$$

Expanding this, we obtain

$$a_1(\xi_{111} - \xi_{122}) + a_2(\xi_{211} - \xi_{222}) + \ldots + a_r(\xi_{r11} - \xi_{r22}) = 0 \quad (2.8)$$

where the coefficients a_i's depend on the correlation model. This expression will be identically equal to zero only if

$$\xi_{111} = \xi_{122}, \xi_{211} = \xi_{222}, \ldots, \xi_{r11} = \xi_{r22}$$

Iterating (2.8), (c-2) times we get (2.6).

Next, we consider

$$\text{Cov}\left(y_{11}y_{12}, \hat{T}_{Qm}\right) - \text{Cov}\left(y_{11}y_{13}, \hat{T}_{Qm}\right) = 0$$

Expanding, we get

$$a_1'(\xi_{112} - \xi_{113}) + a_1''(\xi_{122} - \xi_{133}) + \ldots + a_r'(\xi_{r12} - \xi_{r13}) + a_r''(\xi_{r22} - \xi_{r33}) = 0$$

where a_i' and a''_i, $\forall i$ depend on the correlation model.

Using (2.6), we obtain

$$a_1'(\xi_{112} - \xi_{113}) + \ldots + a_r'(\xi_{r12} - \xi_{r13}) = 0$$

This will be identically equal to zero if and only if

$$\xi_{112} = \xi_{113}, \ldots, \xi_{r12} = \xi_{r13} \quad (2.9)$$

Iterating (2.9), (c(c-1)-2) times we get (2.7).

In view of this we can express $\hat{T}_{Qm}$ as a linear combination of mean of sample values of y_{ik}^2 and $y_{ik}\,y_{il}$; each sample mean is formed from the sample values lying in a row of rectangular sub-pattern.

3. Sampling on Two Occasions

Let (x_i, y_i; i = 1, ..., N) be the values of the characteristics on first and second occasion respectively. A simple random sample (with replacement) of 'n' units is obtained on the first occasion. A random sub-sample of $m = n\lambda$ units is retained for use on the second occasion. An independent sample of $u = n - m = n\mu$ units is selected (unmatched with the first occasion).

Let $\overline{X}, \overline{Y}$ and σ_x^2, σ_y^2 respectively be the population means and variances on the two occasions

where

$$\overline{X} = \frac{1}{N}\sum_k^N x_k ; \overline{Y} = \frac{1}{N}\sum_k^N y_k ; \sigma_x^2 = \frac{1}{N}\sum_k^N \left(x_k - \overline{X}\right)^2 ; \text{ and}$$

$$\sigma_y^2 = \frac{1}{N}\sum_k^N \left(y_k - \overline{Y}\right)^2$$

The parameter of interest is σ_y^2.

In view of Theorem 3 we propose the following estimator of σ_y^2 as

$$\hat{T}_{2y} = \frac{a}{u}\sum_k^u x_k^2 + \frac{b}{u(u-1)}\sum_k^u \sum_{j \neq (k)}^u x_k x_j + \frac{c}{m}\sum_k^m x_k^2 + \frac{d}{m(m-1)}\sum_k^m \sum_{j \neq (k)}^m x_k x_j$$

$$+ \frac{e}{m}\sum_k^m y_k^2 + \frac{f}{m(m-1)}\sum_k^m \sum_{j \neq (k)}^m y_k y_j + \frac{g}{u}\sum_k^u y_k^2 + \frac{h}{u(u-1)}\sum_k^u \sum_{j \neq (k)}^u y_k y_j$$

If we assume that a + b = 0, c + d = 0, e + f = 0 and h + g = 0 i.e. the individual group estimators are unbiased, then

$$E\left(\hat{T}_{2y}\right) = (a + c)\sigma_x^2 + (e + g)\sigma_y^2$$

Further, for unbiasedness of $\hat{T}_{2y}$ it is necessary that a + c = 0 and e + g = 1. Therefore, $\hat{T}_{2y}$ reduces to

$$\hat{T}_{2y} = \frac{a}{u}\sum_{k}^{u} x_k^2 - \frac{a}{u(u-1)}\sum_{k}^{u}\sum_{j\neq(k)}^{u} x_k x_j - \frac{a}{m}\sum_{k}^{m} x_k^2 + \frac{a}{m(m-1)}\sum_{k}^{m}\sum_{j\neq(k)}^{m} x_k x_j$$

$$+\frac{e}{m}\sum_{k}^{m} y_k^2 - \frac{e}{m(m-1)}\sum_{k}^{m}\sum_{j\neq(k)}^{m} y_k y_j + \frac{(1-e)}{u}\sum_{k}^{u} y_k^2 - \frac{(1-e)}{u(u-1)}\sum_{k}^{m}\sum_{j\neq(k)}^{u} y_k y_j \tag{3.1}$$

One distinct advantage of assumption of unbiasedness of individual group estimators is that $\hat{T}_{2y}$ is reduced to a simple form with only two coefficients to be determined.

We determine the coefficients a and e by using Theorem 2

$$\text{Cov}\left[(x'_u A_u x_u - x'_m A_m x_m), \hat{T}_{2ym}\right] = 0$$

$$\text{Cov}\left[(y'_u A_u y_u - y'_m A_m y_m), \hat{T}_{2ym}\right] = 0 \tag{3.2}$$

where x_u is $u \times 1$ column vector given by $[x_1 \quad x_2 \quad \dots \quad x_u]'$

Other terms in (3.2) like x'_m, y'_m etc. can be similarly defined. Also, A_u, A_m are symmetric semi-positive matrices of order $u \times u$, $m \times m$ whose diagonal and off-diagonal elements sum up to zero and -1 respectively.

It can be seen that

$$\text{Cov}\left[\left(\frac{a}{u}\sum_{k}^{u} x_k^2 - \frac{a}{u(u-1)}\sum_{k}^{u}\sum_{j\neq(k)}^{u} x_k x_j\right), \hat{T}_{2ym}\right] = \frac{a}{u}\left[\mu_{4x} - (\sigma_x^2)^2\right] + 0(u^{-2})$$

$$\text{Cov}\left[\left(\frac{a}{m}\sum_{k}^{m} x_k^2 - \frac{a}{m(m-1)}\sum_{k}^{m}\sum_{j\neq(k)}^{m} x_k x_j\right), \hat{T}_{2ym}\right]$$

$$= -\frac{a}{m}\left[\mu_{4x} - (\sigma_x^2)^2\right] + \frac{e}{m}\left[\mu_{2x,2y} - \sigma_x^2\sigma_y^2\right] + 0(m^{-2})$$

$$\text{Cov}\left[\left(\frac{(1-e)}{u}\sum_{k}^{u} y_k^2 - \frac{(1-e)}{u(u-1)}\sum_{k}^{u}\sum_{j\neq(k)}^{u} y_k y_j\right), \hat{T}_{2ym}\right]$$

$$= \frac{(1-e)}{u}\left[\mu_{4y} - (\sigma_y^2)^2\right] + 0(u^{-2})$$

$$\text{and Cov}\left[\left(\frac{e}{m}\sum_{k}^{m} y_k^2 - \frac{e}{m(m-1)}\sum_{k}^{m}\sum_{j\neq(k)}^{m} y_k y_j\right), \hat{T}_{2ym}\right]$$

$$= \frac{e}{m}\left[\mu_{4y} - \left(\sigma_y^2\right)^2\right] - \frac{a}{m}\left[\mu_{2x,2y} - \sigma_x^2\,\sigma_y^2\right] + 0\left(m^{-2}\right)$$

where

$$\mu_{rx,sy} = \frac{1}{N}\sum_{k}^{N}\left(x_k - \overline{X}\right)^r\left(y_k - \overline{Y}\right)^s$$

Solving for a and e we get

$$a = e\frac{\sigma_y^2 k_1}{n\,\sigma_x^2\,(\beta_{2x}-1)^{\mu}}$$

$$e = nm\frac{k'}{\left[n^2\,k' - u^2\,k_1^2\right]}$$

where $k' = (\beta_{2x}-1)(\beta_{2y}-1), k_1 = \rho_{x^2y^2}\sqrt{\beta_{2x}\,\beta_{2y}} - 1$

$$\beta_{2x} = \frac{\mu_{4x}}{\left(\sigma_x^2\right)^2}, \beta_{2y} = \frac{\mu_{4y}}{\left(\sigma_y^2\right)^2}$$

and $\rho_{x^2y^2}$ is the correlation coefficient between x^2 and y^2.

It may be seen that k′ will have positive values for

$$\left(\beta_{2x}, \beta_{2y}\right) > (1,1) \tag{3.3a}$$

Similarly, k_1 will be positive for

$$\left(\beta_{2x}, \beta_{2y}\right) > \frac{1}{\rho_{x^2y^2}} \tag{3.3b}$$

Here it is worthwhile to mention that the conditions (3.3a) and (3.3b) are satisfied for many distributions encountered in practice.

Substituting the values of a and e in (3.1) we get

$$\hat{T}_{2y} = \frac{k'nm}{\left(k'n^2 - u^2 k_1^2\right)}\left[\frac{\sigma_y^2 k_1 \mu}{n\sigma_x^2(\beta_{2x}-1)}\left(\frac{1}{u}\sum_k^u x_k^2 - \frac{1}{u(u-1)}\sum_k^u \sum_{j\neq(k)}^u x_k x_j\right.\right.$$

$$\left.\left. - \frac{1}{m}\sum_k^m x_k^2 + \frac{1}{m(m-1)}\sum_k^m \sum_{j\neq(k)}^m x_k x_j\right) + \frac{1}{m}\sum_k^m y_k^2 - \frac{1}{m(m-1)}\sum_k^m \sum_{j\neq(k)}^m y_k y_j\right]$$

$$+\left(1 - \frac{k'nm}{\left(k'n^2 - u^2 k_1^2\right)}\right)\left(\frac{1}{u}\sum_k^u y_k^2 - \frac{1}{u(u-1)}\sum_k^u \sum_{j\neq(k)}^u y_k y_j\right)$$

The variance of $\hat{T}_{2y}$ is given by

$$V_{min}\left(\hat{T}_{2y}\right) = Cov\left(y_u' A_u y_u, \hat{T}_{2y}\right) = \frac{\left(k' - \mu k_1^2\right)}{\left(k' - \mu^2 k_1^2\right)}\left[\beta_{2y} - 1\right]\frac{\left(\sigma_y^2\right)^2}{n} \tag{3.4}$$

The feasible optimum values of μ and λ in the sense of minimum variance can be seen equal to

$$\mu_{opt} = \frac{k'}{\left(k' + \sqrt{k'^2 - k'k_1^2}\right)}, \lambda_{opt} = \frac{\sqrt{k'^2 - k'k_1^2}}{\left(k' + \sqrt{k'^2 - k'k_1^2}\right)}$$

and the corresponding minimum variance as

$$V_{min}\left(\hat{T}_{2y}\right)_{opt} = \frac{\left(k' + \sqrt{k'^2 - k'k_1^2}\right)}{2k'}\left[\beta_{2y} - 1\right]\frac{\left(\sigma_y^2\right)^2}{n} \tag{3.5}$$

It can be seen that μ and λ will have real values only if $k' > k_1^2$.

If a completely independent sample is taken on the second occasion, then the estimator is given by

$$\hat{T}_{2y}' = \frac{1}{n}\sum_k^n y_k^2 - \frac{1}{n(n-1)}\sum_k^n \sum_{j\neq(k)}^n y_k y_j$$

and the corresponding variance by

$$V\left(\hat{T}'_{2y}\right) \cong \frac{\left(\sigma_y^2\right)^2}{n}\left[\beta_{2y}-1\right]$$

Thus, the percent gain in precision of $\left(\hat{T}_{2y}\right)_{opt}$ over $\hat{T}'_{2y}$ is given by

$$G = \frac{\left(k' - \sqrt{k'^2 - k'k_1^2}\right)}{k' + \sqrt{k'^2 - k'k_1^2}} \times 100 \tag{3.6}$$

It is well known that for the estimation of population mean in case of sampling on two occasions the optimum value of μ is given by

$$\mu'_{opt} = \frac{1}{1+\sqrt{1-\rho^2}}$$

where, ρ is the correlation coefficient between x and y.

Substituting the optimum value of μ in (3.4) we get

$$V\left(\hat{T}''_{2y}\right) = \frac{\left(k' - k_1^2 + k'\sqrt{1-\rho^2}\right)}{\left[\left(2-\rho^2\right)k' + 2k'\sqrt{1-\rho^2} - k_1^2\right]}\left(1+\sqrt{1-\rho^2}\right)\left[\beta_{2y}-1\right]\frac{\left(\sigma_y^2\right)^2}{n}$$

Assuming $\rho_{x^2y^2} = \rho$ then, the % loss in precision of $\hat{T}''_{2y}$ over $\left(\hat{T}_{2y}\right)_{opt}$ is

$$L = \left[\frac{\left(k' + \sqrt{k'^2 - k'k_1^2}\right)}{2k'\left(k' - k_1^2 + k'\sqrt{1-\rho^2}\right)\left(1+\sqrt{1-\rho^2}\right)} \left\{\left(2-\rho^2\right)k' + 2k'\sqrt{1-\rho^2} - k_1^2\right\} - 1\right] \times 100 \tag{3.7}$$

To get an idea about % gain and loss in precision values of β_{2x}, β_{2y} and ρ and substituting in (3.6) and (3.7) respectively. The results are presented in Table 1and 2.

A close perusal of Table 1 reveals that the percent gain in precision for particular values of β_{2x}, β_{2y} first decreases with increase in values of ρ. It reaches the minimum value and increases again with increase in ρ values. The percent

Table 1. Percentage Gain in precision (G) of $\left(\hat{T}_{2y}\right)_{opt}$ over $\hat{T}'_{2y}$

ρ		β_{2x}, β_{2y}					
		2, 2	2, 3	2, 4	3, 3	3, 4	4, 4
0.1	μ	0.6250	0.5418	0.5235	0.5163	0.5092	0.5051
	Gain	25	8.3690	4.6981	3.2658	1.8462	1.0205
0.2	μ	0.5556	0.5174	0.5081	0.5051	0.5020	0.5006
	Gain	11.1111	3.4831	1.6234	1.0205	0.3963	0.1114
0.3	μ	0.5218	0.5045	0.5010	0.5003	0.5000	0.5006
	Gain	4.3561	0.8946	0.1919	0.0626	0.0064	0.1114
0.4	μ	0.5051	0.5000	0.5007	0.5013	0.5031	0.5051
	Gain	1.0205	0.0051	0.1442	0.2513	0.6275	1.0205
0.5	μ	0.5000	0.5032	0.5074	0.5081	0.5117	0.5147
	Gain	0	0.6395	1.4722	1.6133	2.3386	2.9437
0.6	μ	0.5051	0.5146	0.5221	0.5218	0.5269	0.5307
	Gain	1.0205	2.9211	4.4144	4.3561	5.3818	6.1327
0.7	μ	0.5218	0.5368	0.5481	0.5449	0.5514	0.5556
	Gain	4.3561	7.3579	9.6142	8.9820	10.2899	11.1111
0.8	μ	0.5556	0.5765	0.5937	0.5834	0.5915	0.5953
	Gain	11.1111	15.3024	18.732	16.6764	18.2928	19.0569
0.9	μ	0.6250	0.6562	0.6890	0.6550	0.6655	0.6672
	Gain	25	31.2365	37.8021	30.9944	33.1064	33.4323

Table 2. Percentage Loss in precision (L) of $\left(\hat{T}''_{2y}\right)$ over $\left(\hat{T}_{2y}\right)_{opt}$

ρ	μ'_{opt}	β_{2x}, β_{2y}					
		2, 2	2, 3	2, 4	3, 3	3, 4	4, 4
0.1	0.501	–1.1543	–0.0506	–0.0089	–0.0029	–0.0005	–0.0001
0.2	0.505	–0.1008	–0.0020	–0.0001	0	0	0
0.3	0.512	–0.0017	–0.0002	–0.0001	0	0	0
0.4	0.522	–0.0011	0	–0.0003	–0.0004	–0.0009	–0.0011
0.5	0.536	0	–0.0027	–0.0047	–0.0049	–0.0054	–0.0051
0.6	0.556	–0.0103	–0.0191	–0.0191	–0.0192	–0.0169	–0.0145
0.7	0.583	–0.0642	–0.0607	–0.0448	–0.0499	–0.0390	–0.0317
0.8	0.625	–0.2016	–0.1318	–0.0658	–0.1048	–0.0738	–0.0601
0.9	0.696	–0.4713	–0.1810	–0.0072	–0.1908	–0.1118	–0.1011

gain in precision for low values of ρ decreases with the increase in β_{2x}, β_{2y} values while for higher values of ρ it increases with the increase in ρ values. Also, from Table 2, it is clear that the percent loss in precision is negligible for all values of β_{2x}, β_{2y} and ρ. This suggests that, in repeat surveys, the problem of estimation of population variance can be simultaneously considered with the estimation of population mean.

REFERENCES

[1] Cochran, W.G. (1977). *Sampling Techniques* (Third edition). John Wiley and Sons, Inc., New York, London.

[2] Eckler, A.R. (1955). Rotation sampling. *Ann. Math. Satist.*, **26**, 664-685.

[3] Jessen, R.J. (1942). Statistical investigation of a sample survey for obtaining farm facts. *Iowa Agri. Expt. Station Res. Bul.*, **304**.

[4] Patterson, H.D. (1950). Sampling on successive occasions with partial replacements of units. *J. Roy. Statist. Soc.*, **B12**, 241-255.

[5] Raj, D. (1965). On sampling over two occasions with probability proportional to size. *Ann. Math. Statist.*, **36**, 327-330.

[6] Vijayan, K. (1975). On estimating the variance in unequal probability sampling. *Jour. Amer. Statist. Assoc.*, **70**, 713-716.