Journal of the Korean
Statistical Society
Vol. 25, No. 1, 1996

Regression Estimators with Unequal Selection Probabilities on Two Successive Occasions

Kyu Seong Kim [1]

Abstract

In this paper, we propose regression estimators based on a partial replacement sampling scheme over two successive occasions and derive the minimum variances of them. PPSWR, RHC, πPS and PPSWOR schemes are considered to select unequal probability samples on two occasions. Simulation results over four populations are given for comparison of composite estimators and regression estimators.

Key Words : Successive occasion sampling; Unequal probability sampling; Composite estimator; Regression estimator.

[1] Department of Computer Science and Statistics, Seoul City University, Seoul, 130-743, Korea

1. INTRODUCTION

We suppose a sample survey to be conducted on two successive occasions. Assume that a population U consists of N identifiable units and the population units are maintaind until two successive surveys are finished. Let Y_{ik} be the survey value of the unit k at the ith occasion($i = 1, 2, k = 1, ..., N$) and T_i be the population total at the ith occasion. Before the survey we suppose initial selection probabilities, say p_k $(k = 1, ..., N)$, are known for each of the population units.

In estimating the current population total T_2, a partial replacement sampling scheme has been used to increase the efficiency of estimation. This sampling scheme is as follows:

(1) On the first occasion the first sample s_1 of size n is selected from the population by a sampling design $p(.)$.

(2) On the second occasion a matched sample s_m of size $m(\leq n)$ is drawn from this s_1 by a conditional sampling design $p(.|s_1)$ and an unmatched sample s_u of size u is taken from the population independently of s_1 and s_m by $p(.)$. Then the second sample s_2 is composed of these s_m and s_u. Here we assume $m + u = n$.

For this sampling scheme some unequal probability sampling designs have been considered as $p(.)$ and $p(.|s_1)$. Des Raj(1965), Park(1974), and Prasad and Graham(1994) applied probability proportional to size sampling with replacement, briefly PPSWR, and Chotai(1974), Ghangurde and Rao(1969), Park and Lee(1993) used Rao-Hartley-Cochran, briefly RHC, method to this sampling scheme.

Based on the partial replacement samples s_1, s_m, and s_u, most of researchers have used composite estimators for the current total of the form

$$t = \phi t_m + (1 - \phi)t_u \tag{1.1}$$

where t_m is based on Y_{1k}-values in s_1 and Y_{2k}-values in s_m and t_u is based on Y_{2k}-values in s_u, and ϕ is a weight, $0 \leq \phi \leq 1$.

In this paper, after review of composite estmators briefly in section 2, we propose regression estimators and derive minimum variances of them in section 3 and compare regression estimators and composite estimators numerically over four populations by Monte Carlo simulation study in section 4. Finally, conclusions are shown in section 5.

2. COMPOSITE ESTIMATORS

Raj(1965) select PPSWR sample s_1 and s_u from U and draw simple random sub-sample s_m from s_1 and propose a composite estimator for T_2

$$t_1 = \phi(\sum_{k\in s_1} \frac{Y_{1k}}{np_k} + \sum_{k\in s_m} \frac{Y_{2k}-Y_{1k}}{mp_k}) + (1-\phi)\sum_{k\in s_u} \frac{Y_{2k}}{(n-m)p_k}, \qquad (2.1)$$

where ϕ is a weight. We define

$$V_{ij} = \frac{1}{2}\sum_{k\in U}\sum_{l\in U} p_k p_l (\frac{Y_{ik}}{p_k} - \frac{Y_{il}}{p_l})(\frac{Y_{jk}}{p_k} - \frac{Y_{jl}}{p_l}) \quad for\ i,j=1,2 \qquad (2.2)$$

and assume that $V = V_{11} = V_{22}$. Then for a given matching fraction $\lambda(= m/n)$ the minimum variance of t_1 with respect to ϕ is given by

$$V_{min,\lambda}(t_1) = \frac{V}{n}\frac{1+(1-\lambda)(1-2\rho)}{1+(1-\lambda)^2(1-2\rho)}, \qquad (2.3)$$

where $\rho = V_{12}/\sqrt{V_{11}V_{22}}$. Also (2.3) can be minimized with optimun choice of λ by

$$V_{min}(t_1) = \frac{V}{2n}(1+\sqrt{2(1-\rho)}). \qquad (2.4)$$

Chotai(1974) draw s_1, s_m and s_u by RHC method and suggest a composite estimator for T_2

$$t_2 = \phi(\sum_{k\in s_1} \frac{Y_{1k}}{p_k}Q_{1k} + \sum_{k\in s_m} \frac{Y_{2k}-Y_{1k}}{p_k}Q_{2k}) + (1-\phi)\sum_{k\in s_u} \frac{Y_{2k}}{p_k}Q_{3k}, \qquad (2.5)$$

where $Q_{ik}(i=1,2,3)$ are the total p-values of the group containing unit k.

Under the assumption that $V = V_{11} = V_{22}$ and N/n, $N/(n-m)$ and n/m are all integers, the variance of t_2 with optimum value of ϕ is given by

$$V_{min,\lambda}(t_2) = \frac{N-n(1-\lambda)}{(N-1)(1-\lambda)}\frac{V}{n}\{1 - \frac{(N-n(1-\lambda))\lambda}{N-2n\lambda(1-\lambda)+N(1-\lambda)^2(1-2\rho)}\} \qquad (2.6)$$

and the minimum variance of t_2 inserting optimum values of ϕ and λ in (2.6) is obtained as follows

$$V_{min}(t_2) = \frac{V}{2n}\frac{N}{N-1}\{1 - \frac{n}{N} + \sqrt{2(1-\rho)}\}. \tag{2.7}$$

From (2.4) and (2.7) it can be shown that t_2 is more efficient than t_1 for all ρ and n/N, which means that PPSWR scheme can be uniformly improved by RHC scheme on two occasion sampling.

3. REGRESSION ESTIMATORS

We now suggest regression estimators based on the sample taken by the partial replacement sampling scheme above mentioned in section 1. Assume that $t_{11}(t_{1m})$ is an unbiased estimator for T_1 defined on $s_1(s_m)$ and $t_{2m}(t_{2u})$ is unbiased for T_2 on $s_m(s_u)$. In the following theorem we construct a regression estimator and derive the minimum variance of it.

Theorem 3.1. We suppose a partial replacement sampling scheme on two occasions. Then the regression estimator of the form

$$t = \phi(t_{2m} + K(t_{11} - t_{1m})) + (1-\phi)t_{2u}, \tag{3.1}$$

where ϕ is a weight and K is a regression coefficient, is unbiased for T_2 and has minimum variance

$$min_{\phi,K}Var(t) = Var(t_{2u})(1 - \phi_{opt}), \tag{3.2}$$

where

$$\phi_{opt} = \frac{Var(t_{2u})}{Var(t_{2u}) + Var(t_{2m}) - K_{opt}ECov(t_{1m}, t_{2m}|s_1)}, \tag{3.3}$$

$$and \quad K_{opt} = ECov(t_{1m}, t_{2m}|s_1)/EVar(t_{1m}|s_1) \tag{3.4}$$

Here ϕ_{opt} and K_{opt} are optimum values of ϕ and K respectively.

Proof. Because t_{11} and t_{1m} are unbiased for T_1 and t_{2m} and t_{2u} are unbiased for T_2, t is trivially unbiased. Since s_u is selected independently of s_1 and s_m and from the fact that $Cov(t_{11} - t_{1m}, t_{2m}) = -ECov(t_{1m}, t_{2m}|s_1)$ and $Var(t_{11} - t_{1m}) = EVar(t_{1m}|s_1)$, we have

$$\begin{aligned} Var(t) &= \phi^2 Var(t_{2m} + K(t_{11} - t_{1m})) + (1-\phi)^2 Var(t_{2u}) \\ &\geq Var(t_{2u})\{1 - \frac{Var(t_{2u})}{Var(t_{2u}) + Var(t_{2m} + K(t_{11} - t_{1m}))}\} \\ &\geq Var(t_{2u})\{1 - \frac{Var(t_{2u})}{Var(t_{2u}) + Var(t_{2m}) - [ECov(t_{1m}, t_{2m}|s_1)]^2/EVar(t_{1m}|s_1)}\}. \end{aligned}$$

By denoting $K_{opt} = ECov(t_{1m}, t_{2m}|s_1)/EVar(t_{1m}|s_1)$ the proof is complete.

According to the sampling design $p(.)$ and $p(.|s_1)$, the regression estimator (3.1) has different forms. In case the Raj's scheme and Chotai's scheme are adapted, we can suggest new regression estimators for T_2.

Under the Raj's scheme, the regression estimator for T_2 is given by

$$t_3 = \phi(\sum_{k\in s_m} \frac{Y_{2k}}{mp_k} + K(\sum_{k\in s_1} \frac{Y_{1k}}{np_k} - \sum_{k\in s_m} \frac{Y_{1k}}{mp_k})) + (1-\phi)\sum_{k\in s_u} \frac{Y_{2k}}{(n-m)p_k} \quad (3.5)$$

and minimum variance under the assumption $V = V_{11} = V_{22}$ can be obtained as

$$V_{min,\lambda}(t_3) = \frac{V}{n}\frac{1-(1-\lambda)\rho^2}{1-(1-\lambda)^2\rho^2}, \quad (3.6)$$

where λ is a given matching fraction. Also $V_{min,\lambda}(t_3)$ can be minimized by

$$V_{min}(t_3) = \frac{V}{n}\frac{1+\sqrt{1-\rho^2}}{2}. \quad (3.7)$$

with optimum unmatching fraction $(1-\lambda)_{opt} = 1/(1+\sqrt{1-\rho^2})$.

Clearly, $V_{min,\lambda}(t_1) \geq V_{min,\lambda}(t_3)$ for all ρ and λ. In addition, $V_{min}(t_1) \geq V_{min}(t_3)$ for all ρ and equality holds only if $\rho = 1$.

Under the Chotai's scheme, the regression estimator for T_2 is proposed by

$$t_4 = \phi(\sum_{k\in s_m} \frac{Y_{2k}}{p_k}Q_{2k} + K(\sum_{k\in s_1} \frac{Y_{1k}}{p_k}Q_{1k} - \sum_{k\in s_m} \frac{Y_{1k}}{p_k}Q_{2k})) + (1-\phi)\sum_{k\in s_u} \frac{Y_{2k}}{p_k}Q_{3k} \quad (3.8)$$

and minimum variance with respect to ϕ is given by

$$V_{min,\lambda}(t_4) = \frac{N-n(1-\lambda)}{(N-1)(1-\lambda)}\frac{V}{n}\{1 - \frac{(N-n(1-\lambda))\lambda}{N-2n\lambda(1-\lambda)-N(1-\lambda)^2\rho^2}\}. \quad (3.9)$$

In addition, minimum variance with respect to λ is given by

$$V_{min}(t_4) = \frac{V}{2n}\frac{N}{N-1}\{1 - \frac{n}{N} + \sqrt{1-\rho^2}\}. \quad (3.10)$$

Here optimum matching fraction λ_{opt} is the same of PPSWR scheme.

The regression estimators t_3 defined in (3.5) and t_4 defined in (3.8) may be compared at their minimum variances. From (3.7) and (3.10) we get

$$\frac{V_{min}(t_4)}{V_{min}(t_3)} = \frac{N}{N-1}\{1 - \frac{n}{N}\frac{1}{1+\sqrt{1-\rho^2}}\}, \quad (3.11)$$

which is greater than 1 for all n/N and $\rho > 0$. Some values of $V_{min}(t_4)/V_{min}(t_3)$ were computed for selected values of ρ and n/N and are given in Table 3.1. Here $N \approx N-1$ is assumed.

Table 3.1. Values of $V_{min}(t_4)/V_{min}(t_3)$ for some values of n/N and ρ

ρ	$n/N=0.1$	$n/N=0.2$	$n/N=0.3$	$n/N=0.4$	$n/N=0.5$
0.00	0.950	0.900	0.850	0.800	0.750
0.10	0.950	0.900	0.850	0.799	0.749
0.20	0.949	0.899	0.849	0.798	0.747
0.30	0.949	0.898	0.846	0.795	0.744
0.40	0.948	0.896	0.843	0.791	0.739
0.50	0.946	0.893	0.839	0.786	0.732
0.60	0.944	0.889	0.833	0.778	0.732
0.70	0.942	0.883	0.825	0.767	0.708
0.80	0.938	0.875	0.812	0.750	0.688
0.90	0.930	0.861	0.791	0.721	0.650
1.00	0.900	0.800	0.700	0.600	0.500

This table says that the efficiency gain increases as n/N or ρ increases.

We now discuss inclusion probability proportion to size sampling, briefly πPS, with Horvitz-Thompson estimator on two occasions. Taking πPS design and simple random sampling design as $p(.)$ and $p(.|s_1)$ respectively, we propose a new regression estimator such as

$$t_5 = \phi\Big(\sum_{k\in s_m}\frac{Y_{2k}}{\pi_k(s_m)} + K\Big(\sum_{k\in s_1}\frac{Y_{1k}}{\pi_k(s_1)} - \sum_{k\in s_m}\frac{Y_{1k}}{\pi_k(s_m)}\Big)\Big) + (1-\phi)\sum_{k\in s_u}\frac{Y_{2k}}{\pi_k(s_u)}, \tag{3.12}$$

where $\pi_k(s_m)$, $\pi_k(s_1)$ and $\pi_k(s_u)$ are inclusion probabilities of the unit k in s_m, s_1 and s_u respectively. By Theorem 3.1 the minimum variance of t_5 with respect to ϕ and K can be calculated by the following covariance components:

$$Var\Big(\sum_{k\in s_m}\frac{Y_{2k}}{\pi_k(s_m)}\Big) = \frac{1}{2}\sum_{k\in U}\sum_{l\in U}\Big(p_k p_l - \frac{m-1}{n(n-1)m}\pi_{kl}(s_1)\Big)\Big(\frac{Y_{2k}}{p_k} - \frac{Y_{2l}}{p_l}\Big)^2,$$

$$Var\Big(\sum_{k\in s_u}\frac{Y_{ik}}{\pi_k(s_u)}\Big) = \frac{1}{2}\sum_{k\in U}\sum_{l\in U}\Big(p_k p_l - \frac{1}{(n-1)^2}\pi_{kl}(s_u)\Big)\Big(\frac{Y_{2k}}{p_k} - \frac{Y_{2l}}{p_l}\Big)^2, \tag{3.13}$$

$$\text{and } ECov(\sum_{k\in s_m}\frac{Y_{ik}}{\pi_k(s_m)},\sum_{k\in s_m}\frac{Y_{jk}}{\pi_k(s_m)}|s_1)=\frac{1}{2}\frac{n-m}{mn^2(n-1)}\sum_{k\in U}\sum_{l\in U}(\frac{Y_{ik}}{p_k}-\frac{Y_{il}}{p_l})(\frac{Y_{jk}}{p_k}-\frac{Y_{jl}}{p_l})$$
$$for \quad i,j=1,2,$$

where $\pi_{kl}(s_1)=Pr\{k\&l\in s_1\}$ and $\pi_{kl}(s_u)=Pr\{k\&l\in s_u\}$.

Finally, we consider probability proportional to size sampling without replacement ,briefly PPSWOR, with Murthy estimator on two occasions. The first sample s_1 and unmatched sample s_u are taken by PPSWOR and matched sample s_m is drawn by simple random sampling. Then we present another new regression estimator based on these samples,

$$t_6=\phi(\sum_{k\in s_m}\frac{Y_{2k}p(s_1|k)}{p(s_1)}+K(\sum_{k\in s_1}\frac{Y_{1k}p(s_1|k)}{p(s_1)}-\frac{n}{m}\sum_{k\in s_m}\frac{Y_{1k}p(s_1|k)}{p(s_1)}))$$
$$+(1-\phi)\sum_{k\in s_u}\frac{Y_{2k}p(s_u|k)}{p(s_u)}, \tag{3.14}$$

where $p(s|k)$ is the conditional probability of drawing sample s given the unit k has been drawn in the first drawing.

Applying Rao and Vijyan's result(1977), we can obtain the following covariance components:

$$Var(\sum_{k\in s_m}\frac{Y_{2k}p(s_1|k)}{p(s_1)})=\frac{1}{2}\sum_{k\in U}\sum_{l\in U}(1-\sum_{s_1\ni k\&l}\frac{p(s_1|k)p(s_1|l)}{p(s_1)})p_kp_l(\frac{Y_{2k}}{p_k}-\frac{Y_{2l}}{p_l})^2$$
$$+E\{\frac{n-m}{2m(n-1)}\sum_{k\in s_1}\sum_{l\in s_1}(\frac{Y_{2k}p(s_1|k)}{p(s_1)}-\frac{Y_{2l}p(s_1|l)}{p(s_1)})\},$$
$$Var(\sum_{k\in s_u}\frac{Y_{2k}p(s_u|k)}{p(s_u)})=\frac{1}{2}\sum_{k\in U}\sum_{l\in U}(1-\sum_{s_u\ni k\&l}\frac{p(s_u|k)p(s_u|l)}{p(s_u)})p_kp_l(\frac{Y_{2k}}{p_k}-\frac{Y_{2l}}{p_l})^2, \tag{3.15}$$

$$\text{and } ECov(\sum_{k\in s_m}\frac{Y_{ik}p(s_1|k)}{p(s_1)},\sum_{k\in s_m}\frac{Y_{jk}p(s_1|k)}{p(s_1)})$$
$$=E\{\frac{n-m}{2m(n-1)}\sum_{k\in s_1}\sum_{l\in s_1}(\frac{Y_{ik}p(s_1|k)}{p(s_1)}-\frac{Y_{il}p(s_1|l)}{p(s_1)})(\frac{Y_{jk}p(s_1|k)}{p(s_1)}-\frac{Y_{jl}p(s_1|l)}{p(s_1)})\}.$$

Minimum variance of t_6 can be calculated by these covariance components.

4. COMPARISON OF ESTIMATORS

In this section, We compare composite estimators and regression estimators numerically over real and simulated populations. For Monte Carlo simulation study, we will construct feasible estimators of t_1 - t_6 in the form of

$$\widehat{t_i} = \widehat{\phi}_i t_m + (1-\widehat{\phi}_i)t_u, \quad i=1,2 \tag{4.1}$$
$$and \;\; \widehat{t_i} = \widehat{\phi}_i(t_{2m} + \widehat{K}_i(t_{11} - t_{1m})) + (1-\widehat{\phi}_i)t_{2u}, \quad i=3,...,6,$$

where $\widehat{\phi}_i$ and $\widehat{K}_i$ are estimators based on the partial replacement samples. To estimate V_{ij} in (2.2) we define two quantities

$$b_{ij} = \frac{1}{2m(m-1)} \sum_{k\in s_m}\sum_{l\in s_m}(\frac{Y_{ik}}{p_k} - \frac{Y_{il}}{p_l})(\frac{Y_{jk}}{p_k} - \frac{Y_{jl}}{p_l}), \quad for \;\; i,j=1,2 \tag{4.2}$$

$$and \; w_{ij} = \frac{(N-1)m}{N(m-1)}\{\sum_{k\in s_m}\frac{Y_{ik}Y_{jk}}{p_k^2}Q_{2k} - \sum_{k\in s_m}\frac{Y_{ik}}{p_k}Q_{2k}\sum_{k\in s_m}\frac{Y_{jk}}{p_k}Q_{2k}\}, \quad i,j=1,2. \tag{4.3}$$

Then b_{ij} and w_{ij} become unbaised estimators for V_{ij} in PPSWR and RHC schemes respectively.

For $\widehat{t}_1$ and $\widehat{t}_3$, optimum weight and optimum regression coefficient can be estimated by means of b_{ij} as

$$\widehat{\phi}_1 = \{\frac{n}{m} + \frac{(n-m)^2}{mn}(\frac{b_{11}}{b_{22}} - 2\frac{b_{12}}{b_{22}})\}^{-1}, \tag{4.4}$$

$$\widehat{\phi}_3 = \frac{mn}{n^2 - (n-m)^2\widehat{\rho}_3^2}, \quad \widehat{\rho}_3^2 = \frac{b_{12}^2}{b_{11}b_{22}}, \tag{4.5}$$

$$and \;\; \widehat{K}_3 = \frac{b_{12}}{b_{11}}. \tag{4.6}$$

Also for $\widehat{t}_2$ and $\widehat{t}_4$ in RHC scheme, we have

$$\widehat{\phi}_2 = \{1 + \frac{(N-m)(n-m)}{m(N-n+m)} + \frac{N(n-m)^2}{nm(N-n+m)}(\frac{w_{11}}{w_{22}} - 2\frac{w_{12}}{w_{22}})\}^{-1}, \tag{4.7}$$

$$\widehat{\phi}_4 = \frac{(N-n+m)mn}{n(Nn-2mn+2m^2) - N(n-m)^2\widehat{\rho}_4^2}, \quad \widehat{\rho}_4^2 = \frac{w_{12}^2}{w_{11}w_{22}}, \tag{4.8}$$

$$\text{and} \quad \widehat{K}_4 = \frac{w_{12}}{w_{11}}. \tag{4.9}$$

For $\hat{t}_5$ and $\hat{t}_6$, after estimating $Var(t_{2m}), Var(t_{2u})$ and $ECov(t_{1m}, t_{2m}|s_1)$ unbiasedly, we can construct $\widehat{\phi}$ and $\widehat{K}$ by substituting these unbaised covariance estimates for covariances in ϕ and K.

In order to examine numerical comparisons, 4 populations are considered. The first population is populations of 32 municipalities in Sweden. Survey variable Y_1 is the population in 1975 (in thousands) and Y_2 is the population in 1985(in thousands). The population 2-4 are simulated as follows,

First, we choose Y_1 from normal distribution with mean 10 and variance 25,

$$Y_{1k} \sim N(0, 25), \quad k = 1, ..., 32$$

and size measure p_k is generated based on Y_{1k},

$$p_k \propto Y_{1k} + \epsilon_k, \quad \epsilon_k \sim N(0,1), \quad k = 1, ..., 32, \quad \sum_{k=1}^{32} p_k = 1.$$

For the survey values on the second occasion we set

$$Y_{ik} \propto \alpha + \beta Y_{ik} + \epsilon_k, \quad \epsilon_k \sim N(0,1), \quad k = 1, ..., 32, \quad i = 2, 3, 4$$

and we take α and β for three populations such as

population 2 : α=0, β=0.5,
population 3 : α=0, β=2.5,
population 4 : α=2, β=1.0.

Finally we normalize $Y_{ik} (i = 2, 3, 4)$ to have similar variances. Basic characteristics of these populations are given in Table 4.1.

Table 4.1. Basic characteristics for 4 populations.

population	survey variable	total	variance	ρ
1	Y_1	1162.0	270.80	0.888
	Y_2	1286.0	309.19	
2	Y_1	336.2	5.65	0.449
	Y_2	181.6	2.96	
3	Y_1	336.2	5.65	0.928
	Y_2	275.6	4.28	
4	Y_1	336.2	5.65	0.834
	Y_2	267.6	3.52	

Table 4.2. Comparison of feasible estimators $\widehat{t}_1$ - $\widehat{t}_6$ over 4 populations by 1,000 replications.

populations			1		2		3		4	
est.	sample size		ave.	var.	ave.	var.	ave.	var.	ave.	var.
$\widehat{t}_1$	n=8	m=3	1286.8	3367.3	180.9	227.2	275.5	111.8	267.3	129.3
		m=4	1284.6	3258.1	181.2	215.9	275.6	119.4	267.2	130.5
		m=5	1287.1	3387.8	181.4	219.1	275.1	123.9	267.4	131.9
		m=6	1286.1	3597.9	180.0	217.7	275.2	129.1	267.3	134.3
$\widehat{t}_2$	n=8	m=4	1285.3	3213.2	181.5	209.6	274.3	117.7	266.1	128.5
$\widehat{t}_3$	n=8	m=3	1287.7	3199.3	181.0	243.6	275.6	109.8	267.3	126.9
		m=4	1284.5	3246.4	181.3	208.2	275.6	120.2	267.2	130.2
		m=5	1287.0	3405.0	181.4	212.9	275.1	133.0	267.4	136.1
		m=6	1286.1	3794.3	180.9	220.4	275.2	171.8	267.3	170.2
$\widehat{t}_4$	n=8	m=4	1286.7	3174.5	179.6	206.2	274.5	113.6	267.7	128.2
$\widehat{t}_5$	n=8	m=3	1285.7	2461.8	180.9	189.9	275.0	91.3	267.1	105.5
		m=4	1283.0	2452.8	182.0	181.4	275.4	94.5	267.7	105.7
		m=5	1285.2	2589.8	180.7	195.3	275.3	104.5	267.5	109.9
		m=6	1285.6	2721.4	180.6	197.9	274.9	118.4	267.5	124.4
$\widehat{t}_6$	n=8	m=3	1285.5	2497.4	180.5	189.1	275.0	93.4	267.3	108.3
		m=4	1284.3	2461.4	181.1	180.2	275.1	95.6	267.5	106.5
		m=5	1283.7	2611.1	180.1	194.2	275.3	106.6	267.4	108.1
		m=6	1285.5	2699.4	180.5	196.4	274.3	117.4	267.3	120.6

n : sample size of the first sample s_1.
m : sample size of the matched sample s_m.
ave. : sample mean by 1,000 replications.
var. : sample variance by 1,000 replications.

From these populations, we select 8 units for the first sample s_1 by PPSWR, RHC, πPS and PPSWOR schemes. Here, Sampford's rejective method (1967) is taken as πPS scheme and approximated values of π_{kl} suggested by Asok and Sukhatme(1976) are used for π_{kl}. For the second sample s_2, we select matched sample s_m with sample size m=3,4,5 and 6, and choose unmatched sample s_u with size $n - m$ corresponding to matched sample s_m. In RHC scheme, we set $m = 4$ to make n/m and $N/(n-m)$ intergers.

Based on these samples, we construct feasible estimators $\widehat{t}_1 - \widehat{t}_6$ in (4.1) and repeat this process 1,000 times and compute averages and variances of estimates. The results are given in Table 4.2.

On the whole, regression estimators $\widehat{t}_3 - \widehat{t}_6$ yield better results in comparison with composite estimators $\widehat{t}_1$ and $\widehat{t}_2$. Among regression estimators, $\widehat{t}_5$ based on πPS scheme and $\widehat{t}_6$ on PPSWOR scheme have smaller variances than $\widehat{t}_3$ on PPSWR scheme and $\widehat{t}_4$ on RHC's scheme. In addition, $\widehat{t}_6$ needs much more long computing time than $\widehat{t}_5$ in obtaining estimates.

In the first population, optimum proportion of matched sample size m is 4 except the case of $\widehat{t}_3$. But optimum m is shown to be 3 in population 3. This is explaned by correlation coefficient. Since correlation coefficient over population 3 is greater, population 3 tends to have smaller m. In cases of population 2 and 4, similar results are obtained.

5. CONCLUSIONS

We have suggested regression estimators on two occasions and have considered PPSWR, RHC, πPS and PPSWOR schemes in selecting partial replacement samples. By Monte Carlo simulation study, we have compared regression estimators and composite estimators numerically over four populations. In this study, we could find that regression estimators are more effective than composite estimators and confirm that πPS and PPSWOR schemes are well adapted with regression estimators on two occasion sampling.

Under the assumption of the same variances on occasion 1 and 2, the optimum proportion of matched sample was derived in PPSWR and RHC schemes for regression estimators. On the whole, the optimum proportion of matched sample size is smaller as correlation coefficient is larger.

Finally, some problems unsolved in this paper are listed for future studies. First, other sampling schemes, where unmatched sample is selected from complement of the first sample so that these schemes are more effective than our schemes, may be established. Second, optimum proportion of matched sample size over πPS and PPSWOR schemes may be found exactly or approximately.

REFERENCES

(1) Asok,C. and Sukhatme, B. V.(1976). On Sampford's procedure of unequal probability sampling without replacement. *Journal of the American Statistical Association*, **71**, 912-918.

(2) Avadhani, M. S. and Sukhatme, B. V. (1970). A comparison of two sampling procedures with an applicasion to successive sampling. *Journal of the Royal Statistical Society*, **C 19**, 251-259.

(3) Avadhani, M. S. and Sukhatme, B. V. (1972). Sampling on several successive occasions with equal and unequal probabilities and without replacement. *The Australian Journal of Statistics*, **14**, 109-119.

(4) Chotai, J. (1974). A note on the Rao-Hartley-Cochran method for pps sampling over two occasions. *Sankhyā*, **C 36**, 173-180.

(5) Des Raj (1965). On sampling over two occasions with probability proportionate to size. *The Annals of Mathematical Statistics*, **36**, 327-330.

(6) Ghangurde, P. D. and Rao, J. N. K. (1969). Some results on sampling over two occasions. *Sankhyā*, **A 31**, 463-472.

(7) Murthy, M. N. (1957). Ordered and unordered estimators in sampling without replacement. *Sankhyā*, **18**, 379-390.

(8) Park, H. N. (1974). Rotation sampling in time series. *Journal of the Korean Statistical Society*, **2**, 17-23.

(9) Park, H. N., and Lee, K. O. (1993). A study on unequal probability sampling over two successive occasions in time series. *The Korean Journal of Applied Statistics*, **6**, 145-162.

(10) Patterson, H. D. (1950). Sampling on successive occasions with partial replacement of units. *Journal of the Royal Statistical Society*, **B 12**, 241-255.

(11) Prasad, N. G. N. and Graham, J. E. (1994). PPS sampling over two occasions. *Survey Methodology*, **20**, 59-64.

(12) Rao, J. N. K. (1963). On three procedures of unequal probability sampling without replacement. *Journal of the American Statistical Association*, **58**, 202-215.

(13) Rao, J. N. K. and Vijyan, K. (1977). On estimating the variance in sampling with probability proportional to aggregate size. *Journal of the American Statistical Association*, **72**, 579-584.

(14) Sampford, M. R. (1967). On sampling without replacement with unequal probabilities of selection. *Biometrika*, **54**, 499-513.