Optimized Whole Sample Procedures (OWSP) versus Traditional Draw-by-Draw Procedures (TDDP)

Sun-Woong Kim

Steven G. Heeringa

Sung-Joon Hong

Peter W. Solenberger

Dongguk University

&

Survey Research Center, University of Michigan

Overview

- Classification of Procedures for Sampling Without Replacement
- TDDP
- OWSP
- The Elements of the Comparison of TDDP and OWSP
- The Structures of Sample Selection Probabilities
- Estimation of the Superpopulation Model
- Empirical Results
- Concluding Remarks

Classification of Procedures for Sampling Without Replacement

- The classification by manner of selection (Brewer and Hanif (1983)):

  - ✓ Draw-by-Draw Procedures (DDP)
  - ✓ Whole Sample Procedures (WSP)

    Systematic Procedures

    Rejection Procedures

    Other Selection Procedures

- This study gives attention to the comparison between traditional DDP (TDDP) and optimized WSP (OWSP).

TDDP (Traditional DDP)

- Probabilities of selection are defined for each draw and always depend on the units already selected, since the selection is without replacement.

- A number of selection procedures have been developed. Several can be easily run in software such as SAS or SPSS.

- It seems that some of inclusion probabilities proportional to size (πPS) sampling methods have been especially attractive to survey samplers.

- Many books regarding survey sampling basically refer to πPS sampling methods of Mizuno (1952) and Brewer (1963).

- Rao and Bayless (1969) and Bayless and Rao (1970) present the superiority of Murthy (1957)'s method, which is not a πPS sampling method, to some chosen methods.

OWSP (Optimized WSP)

- "Optimized" indicates "variance minimization." One of the oldest methods for variance minimization is that of Raj (1956), which is a πPS sampling method, followed by the one of Jessen (1969).

- The units are not drawn individually and the selection probability for each possible sample of n distinct units is specified. Accordingly, one selection using these probabilities selects the whole sample.

- Kim, Heeringa, and Solenberger (2006) developed a theory of model-based πPS sampling methods for minimizing the model expectation of the variance of the Horvitz and Thompson (H-T) (1952) estimator under superpopulation models.

- The applications on the theory are presented below.

The Elements of the Comparison of TDDP and OWSP

- We compare the sample selection probabilities and the efficiencies of the following methods:

  TDDP: Mizuno (1952), Brewer (1963), and Murthy (1957)

  OWSP: Kim, Heeringa, and Solenberger (2006)

- The comparison focuses on only the case of sample size $n = 2$ (or stratified designs with 2 selections per stratum), which requires a simple sampling procedure and is the most important case in the selection of the primary sampling units (PSUs) in practical multi-stage designs.

- For pPS sampling to select small samples, we would prefer the H-T (1952) estimator or other unbiased estimators to the generalized regression (GREG) estimator, due to the bias.

- For Murthy (1957)'s method, Murthy's estimator, which is an unbiased estimator, can be used.

The Structures of Sample Selection Probabilities

(Notation)

$p_i = x_i / X$: the relative size of the unit, where $X = \sum x_i$ and x_i is an auxiliary variable correlated with the variable of the interest, y_i.

$p(s)$: the selection probability of a sample s

$p_i = \sum_{i \in s} p(s)$: the first-order inclusion probabilities

$p_{ij} = \sum_{i,\, j \in s} p(s)$: the second-order inclusion probabilities, often called the joint probabilities

■ Mizuno (1952)'s Method

- Selection Procedure:

  (1) Select a unit with unequal probabilities, p_i.

  (2) Select the remaining $n-1$ units according to a simple random sampling without replacement.

- Sample Selection Probabilities for $n=2$:

$$p(s)=p_{ij} = p_i \frac{1}{N-1} + p_j \frac{1}{N-1}$$

$$= \frac{1}{X(N-1)} f(x_i, x_j), \text{ where } f(x_i, x_j) = x_i + x_j$$

It can be re-expressed as

$$p(s) = a + b\sum_{i \in s} x_i \text{, where } a = 0 \text{ and } b = \frac{1}{X(N-1)}$$

■ Brewer (1963)'s Method

This method is for the cases where the sample size is just two.

- Selection Procedure:

(1) Draw the first unit with probabilities $\dfrac{p_i(1-p_i)}{1-2p_i}$.

(2) Draw the second unit with probabilities $\dfrac{p_i}{1-p_j}$, where j is the unit drawn first.

- Sample Selection Probabilities for $n=2$:

$$p(s)=p_{ij}=\frac{1}{DX}x_i x_j\left(\frac{1}{X-2x_i}+\frac{1}{X-2x_j}\right), \text{ where } D=\frac{1}{2}\left(1+\sum_{i=1}^{N}\frac{x_i}{X-2x_i}\right)$$

Note that the selection probabilities are the form of

$$p(s) = p_{ij} = \frac{1}{DX} g(x_i, x_j),$$

where $D = \frac{1}{2}\left(1 + \sum_{i=1}^{N} \frac{x_i}{X - 2x_i}\right)$ and $g(x_i, x_j) = x_i x_j \left(\frac{1}{X - 2x_i} + \frac{1}{X - 2x_j}\right)$

■ Murthy's method

- Selection Method:

  The successive units are drawn with probabilities p_i, $\dfrac{p_j}{1-p_i}$, $\dfrac{p_k}{1-p_i-p_j}$, and so on.

- Sample Selection Probabilities for $n=2$:

  $p(s)=\dfrac{1}{X}h(x_i,x_j)$, where $h(x_i,x_j)=x_i x_j\left(\dfrac{1}{X-x_i}+\dfrac{1}{X-x_j}\right)$

  which is similar to that for Brewer's method

■ Methods of Kim, Heeringa, and Solenberger (2006)

- Basic Concepts:

Assume that the finite population is a random sample drawn from a larger population called a superpopulation with the distribution x

If we know or can estimate the superpopulation model at the design stage, it may be useful for selecting samples and give increased precision

- Sample Selection Probabilities for $n=2$ (or $n_h=2$ for strata h=1,…,H):

The probabilities depend on the optimization problems for minimizing the model expectations of the variance of the H-T estimator under a given superpopulation model. Accordingly,

$p(s)$, which is equal to p_{ij}, is an unknown function of x_i and x_j

- Superpopulation Model

$$y_i = a + b x_i + e_i,$$

where $E_x(e_i) = 0$, $V_x(e_i) = s^2 x_i^g \, (g \geq 0)$, and $E_x(e_i e_j) = 0$

Here E_x denotes the model expectation over all the finite populations that can be drawn from the superpopulation.

- Model Expectations (ME) of the Variance of the H-T estimator

(1) Under a form of the variance of the H-T estimator

$$Var_I\left(\hat{Y}_{HT}\right) = \sum_{i=1}^{N} \frac{y_i^2(1-\pi_i)}{\pi_i} + 2\sum_{i=1}^{N}\sum_{j>i}^{N} \frac{\pi_{ij}}{\pi_i \pi_j} y_i y_j - 2\sum_{i=1}^{N}\sum_{j>i}^{N} y_i y_j,$$

the ME is given by

$$E_x\left(Var_I\left(\hat{Y}_{HT}\right)\right) = \frac{X^2}{n}\left[\frac{2a}{n}\sum_{i=1}^{N}\sum_{j>i}^{N}\frac{a+b(x_i+x_j)}{x_i x_j}\pi_{ij} + b^2(n-1)\right]$$

$$+ \sum_{i=1}^{N}\left(X/nx_i - 1\right)\left(\sigma^2 x_i^g + a^2 + b^2 x_i^2 + 2ab\,x_i\right)$$

$$- 2\sum_{i}^{N}\sum_{j>i}^{N}\left(a^2 + ab(x_i + x_j) + b^2 x_i x_j\right)$$

(2) With regard to a different form of the variance of the H-T estimator

$$Var_{II}\left(\hat{Y}_{HT}\right) = \sum_{i=1}^{N}\sum_{j>i}^{N}\left(\pi_i\pi_j - \pi_{ij}\right)\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2,$$

the ME is given by

$$E_x\left(Var_{II}\left(\hat{Y}_{HT}\right)\right) = \frac{\sigma^2 X^{\gamma}}{n}\sum_{i=1}^{N}(1 - n\pi_i)\pi_i^{\gamma-1}$$

$$+ \; 2\alpha\left(\sum_{i=1}^{N}\sum_{j>i}^{N}\left(x_j - x_i\right)\left(\alpha x_i^{-1} + \beta\right)\right)$$

$$+ \; \frac{2\alpha X^2}{n^2}\sum_{i=1}^{N}\sum_{j>i}^{N}\left(x_j^{-1} - x_i^{-1}\right)\left(\alpha x_i^{-1} + \beta\right)\pi_{ij},$$

- Optimization Problems (OP) for Minimizing Model Expectations

OP I:

$$\textbf{\textit{Minimize}} \quad \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{a + b(x_i + x_j)}{x_i x_j} p(s)$$

OP II:

$$\textbf{\textit{Minimize}} \quad \sum_{i=1}^{N} \sum_{j>i}^{N} \left(x_j^{-1} - x_i^{-1}\right)\left(a x_i^{-1} + b\right) p(s)$$

subject to $p_i = \sum_{i \in s} p(s), \; i = 1, \cdots, N$

$cp_i p_j \leq p(s) \leq p_i p_j$, where c is a real number between 0 and 1

Note that both optimization problems depend only on a and b, regardless of the values of s^2 or g in the superpopulation model.

Estimation of the Superpopulation model

As noted by Godfrey, Roshwalb and Wright (1984) and Särndal and Wright (1984), the Harvey (1976) algorithm may be used to calculate the maximum likelihood estimates of a, b, s^2, and g in the superpopulation model.

Harvey’s algorithm is useful for the models in which the variance of the disturbance term is proportional to the auxiliary variable raised to a certain power, which is written by

$$s_i^2 = s^2 x_i^g$$

In his algorithm the starting values of a and b are the ordinary least squares (OLS) estimates and in each iteration the values of a and b depend on s^2 and g, or the reverse.

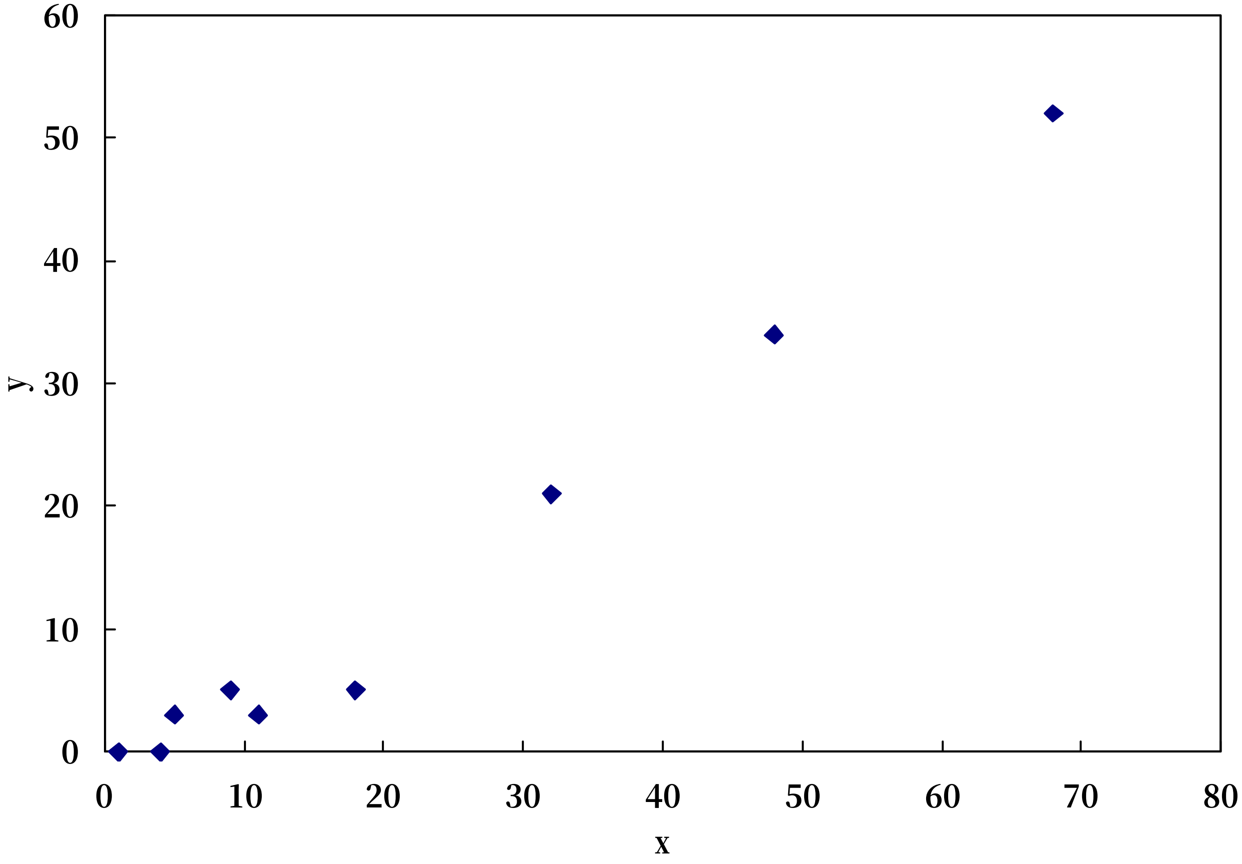
Empirical Results

■ Description and Scatter Plots of Populations

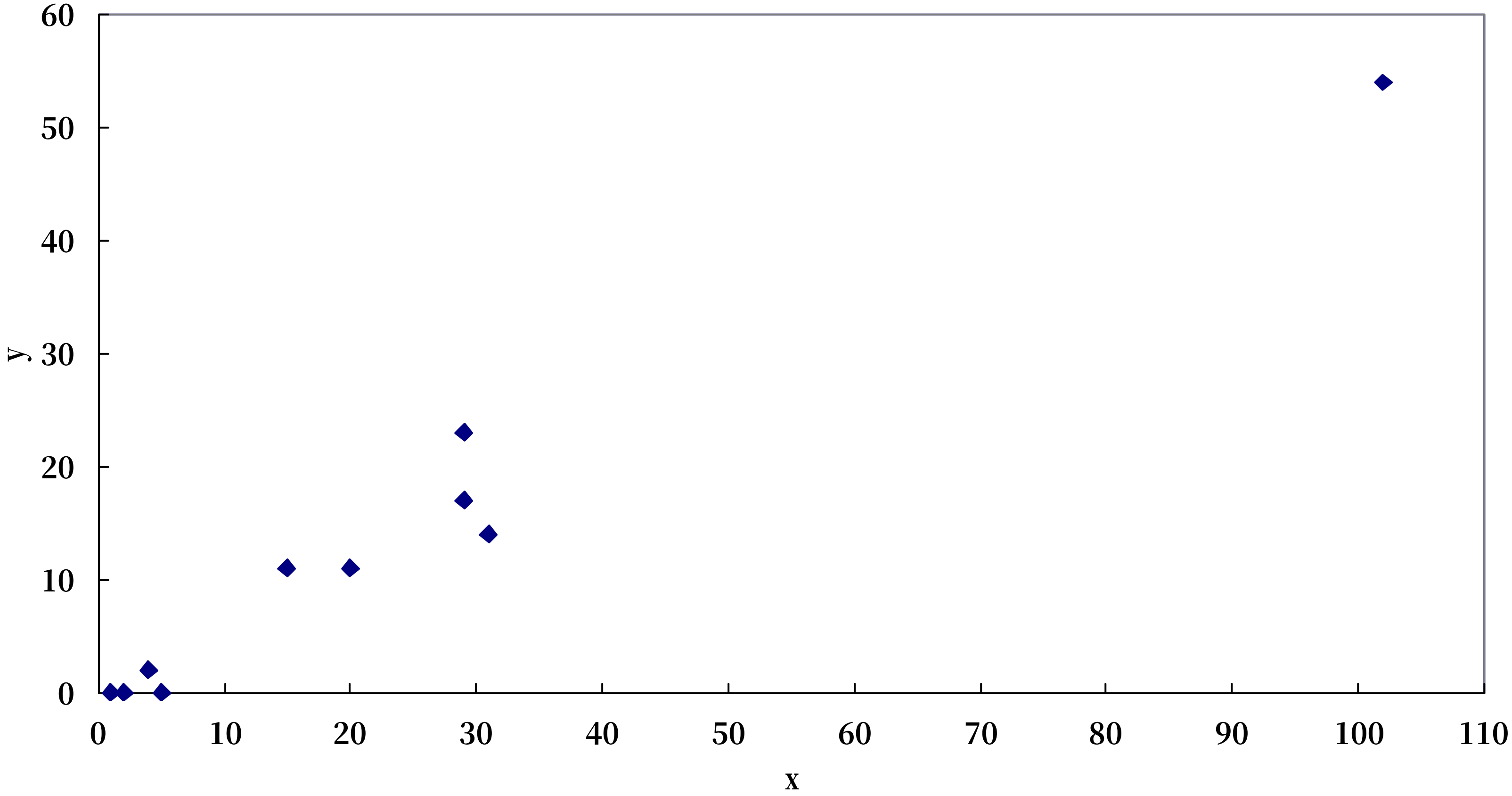
5 Natural Populations used by Rao and Bayless (1969)

No	Source	y	x	N
(1)	Kish (1965)	No. of rented dwelling units	Total no. of dwelling units	10
(2)	Kish (1965)	No. of rented dwelling units	Total no. of dwelling units	10
(3)	Rao (1963)	Corn acreage in 1960	Corn acreage in 1958	14
(4)	Hanurav (1967)	Population in 1967	Population in 1957	20
(5)	Horvitz and Thompson (1952)	No. of Households	Eye-estimated no. of Households	20

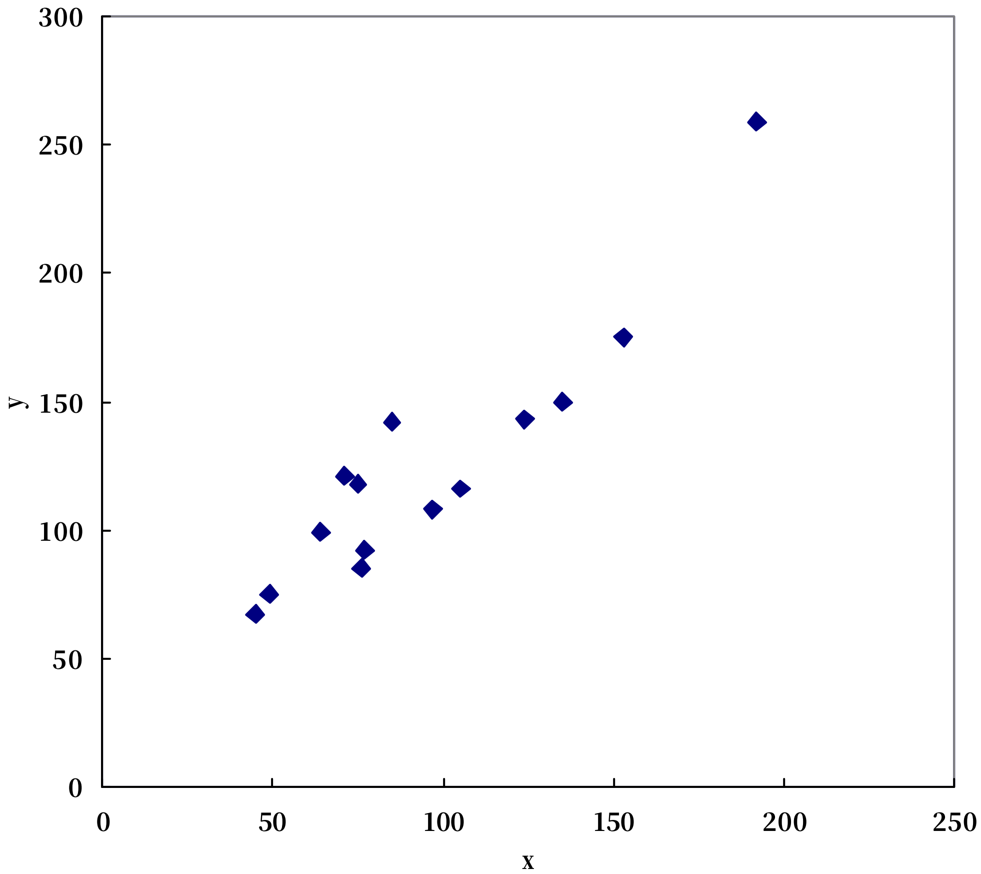
(1) Kish (1965)



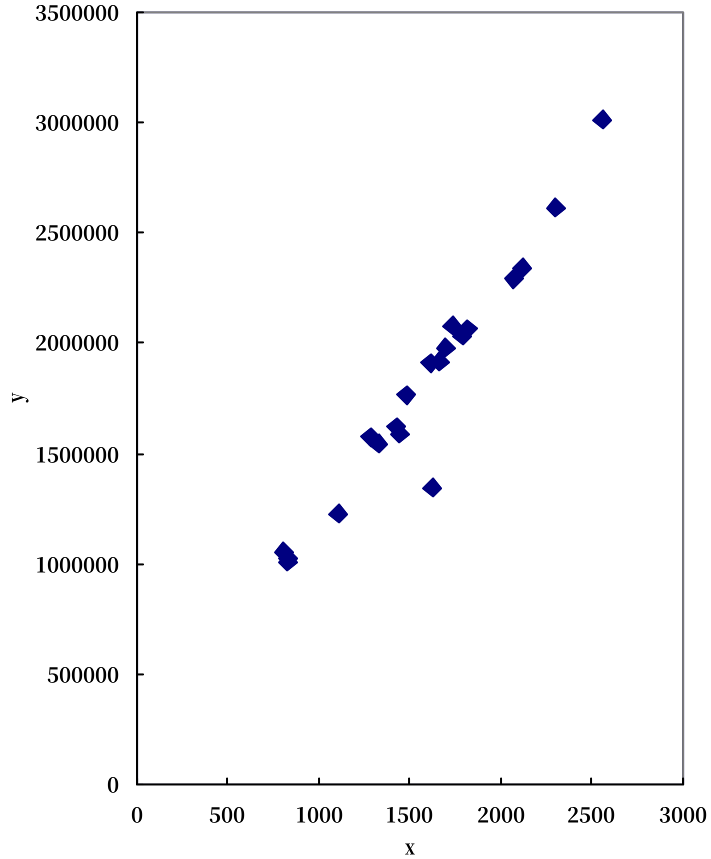
(2) Kish (1965)



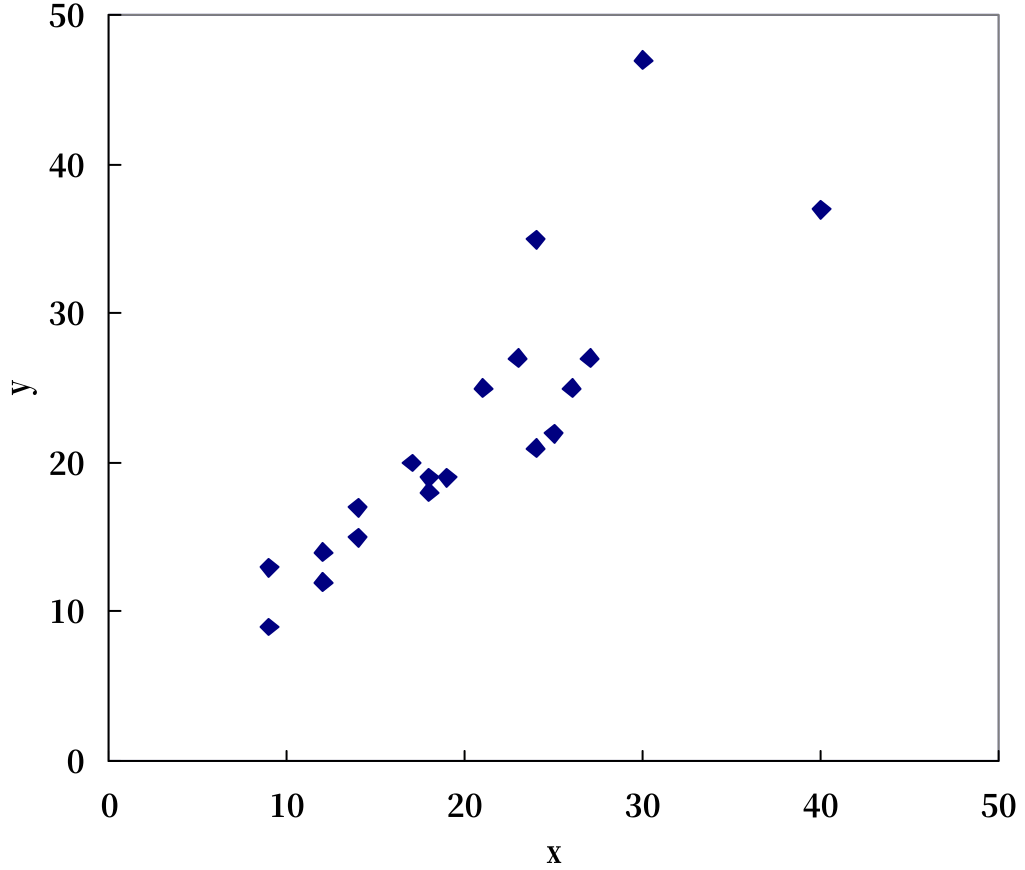
(3) Rao (1963)



(4) Hanurav (1967)



(5) Horvitz and Thompson (1952)



■ Maximum Likelihood Estimates by Harvey (1976)'s algorithm

Pop	Iterations	a	b	s^2	g
(1)	41	-0.5256	0.5058	0.0203	2.2964
(2)	13	-0.8130	0.5951	0.1322	1.4108
(3)	11	25.9264	1.0272	0.1989	1.6099
(4)	38	185718.57	1015.3631	0.5370	3.2817
(5)	14	1.1426	1.0381	0.0038	2.7461

■ The Comparison of the Efficiencies

(1) Kish (1965)

$N = 10$

Method			$Var(\hat{Y})$
Mizuno			2908.04
Brewer			621.98
Murthy			597.57
c	0.1	OP I	794.56
	0.2		792.68
	0.3		621.05
	0.4		659.77
	0.5		NA
c	0.1	OP II	777.96
	0.2		724.61
	0.3		661.91
	0.4		693.59
	0.5		NA

(2) Kish (1965)

$N = 10$

Method			$Var(\hat{Y})$
Mizuno			2183.23
Brewer			567.77
Murthy			480.07
c	0.1	OP I	519.94
	0.2		538.56
	0.3		NA
	0.4		NA
	0.5		NA
c	0.1	OP II	611.75
	0.2		602.16
	0.3		NA
	0.4		NA
	0.5		NA

(3) Rao (1963)

$N = 14$

Method			$Var(\hat{Y})$
Mizuno			53353.35
Brewer			37211.20
Murthy			36771.14
c	0.1	OP I	37536.48
	0.2		32021.02
	0.3		38638.66
	0.4		37588.08
	0.5		38695.92
c	0.1	OP II	53565.07
	0.2		52609.38
	0.3		50038.54
	0.4		44699.52
	0.5		38324.35

(4) Hanurav (1967)

$N = 20$

Method			$Var(\hat{Y})$
Mizuno			1.49 E13
Brewer			3.46 E12
Murthy			3.44 E12
c	0.1	OP I	3.49 E12
	0.2		3.47 E12
	0.3		3.49 E12
	0.4		3.57 E12
	0.5		3.43 E12
c	0.1	OP II	3.33 E12
	0.2		3.48 E12
	0.3		3.46 E12
	0.4		3.50 E12
	0.5		3.49 E12

(5) Horvitz and Thompson (1952)

$N = 20$

Method			$Var(\hat{Y})$
Mizuno			6410.58
Brewer			3010.59
Murthy			3031.40
c	0.1	OP I	2988.96
	0.2		2874.78
	0.3		2831.16
	0.4		2859.07
	0.5		3058.17
c	0.1	OP II	3195.26
	0.2		2962.53
	0.3		3139.18
	0.4		3055.10
	0.5		3009.42

■ The Closeness between $p_i p_j$ and p_{ij}

Here we present several graphs for the example population of (4) Hanurav (1967).There are 190 possible samples for the population.

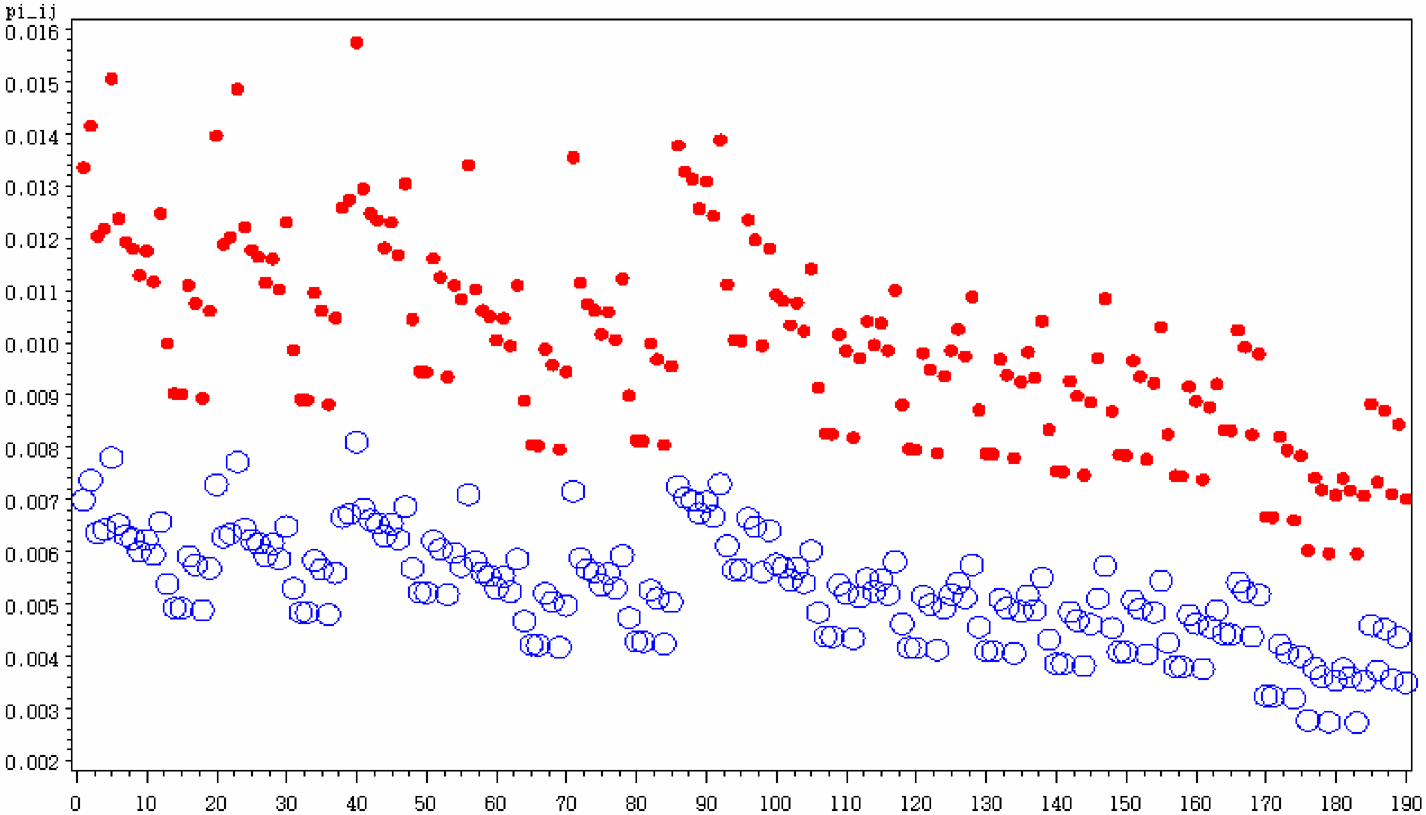
Each scale on the horizontal axis represents an indexed sample, s for s=1,…,190). The scale on the vertical axis represents the values of $p_i p_j$ or p_{ij}.

The dots in the graphs indicate the values of $p_i p_j$ and the circles represent values of the p_{ij}.

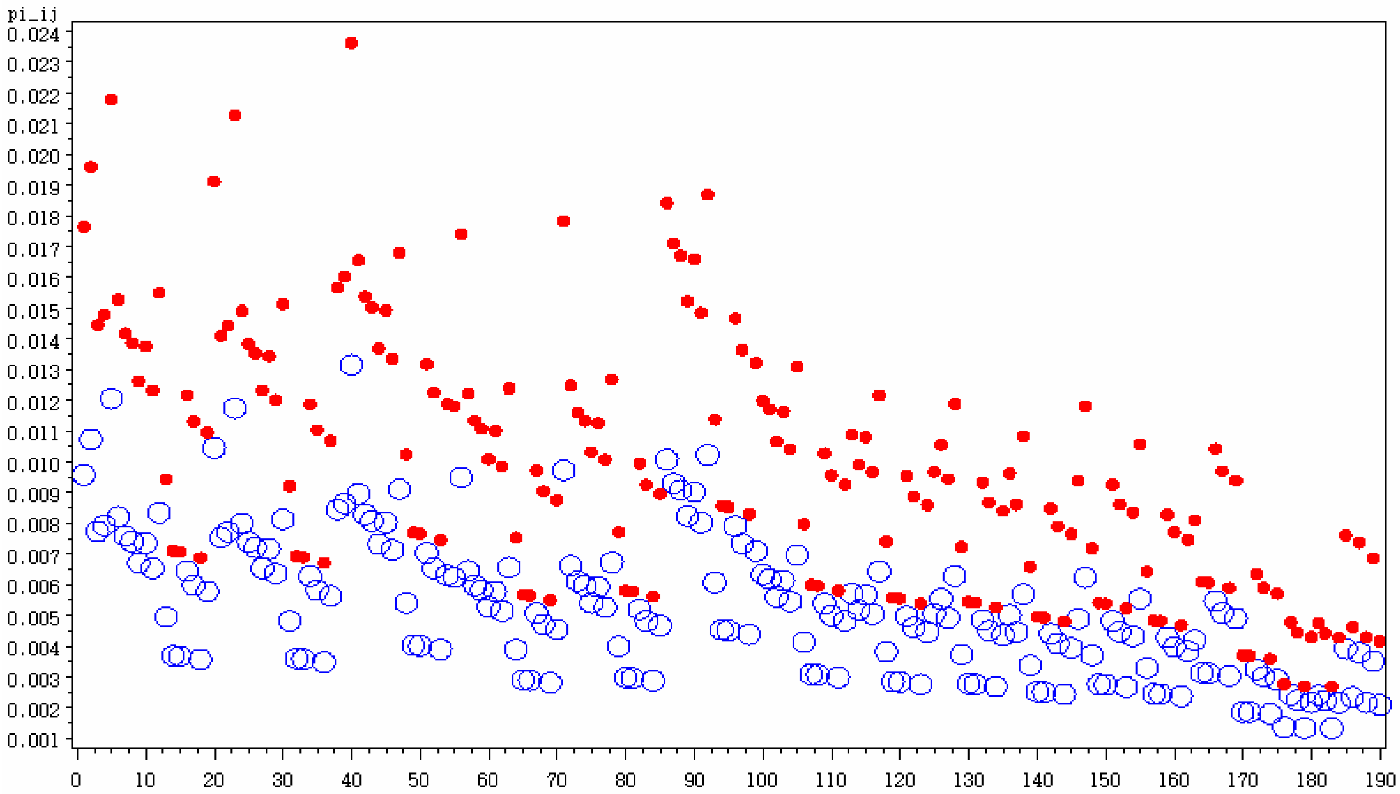
The smaller the vertical distance between the dots and the circles, the smaller variance.

It seems that OP I yields the smallest differences between the values of $p_i p_j$ and p_{ij}

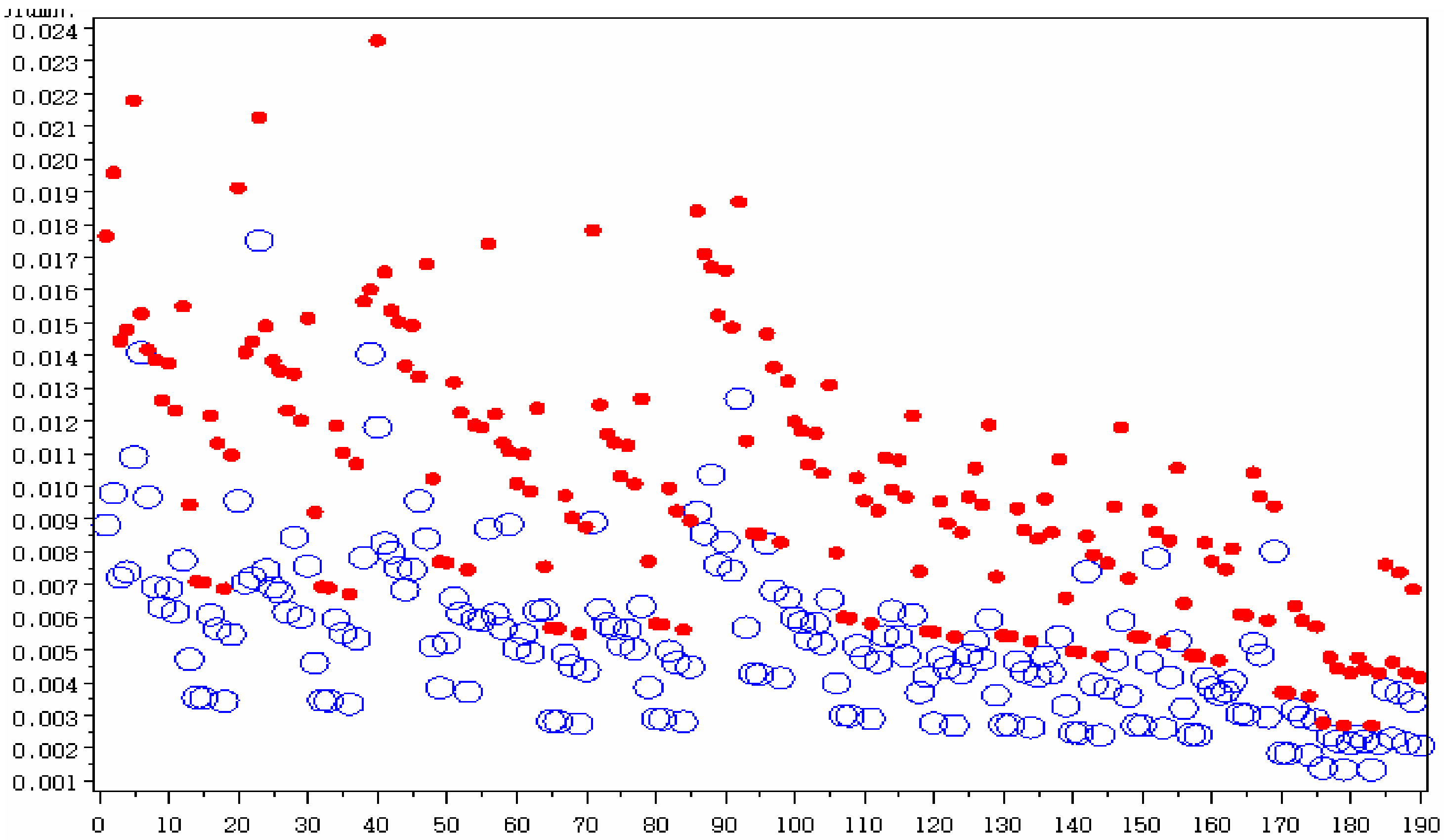
(Mizuno's method)



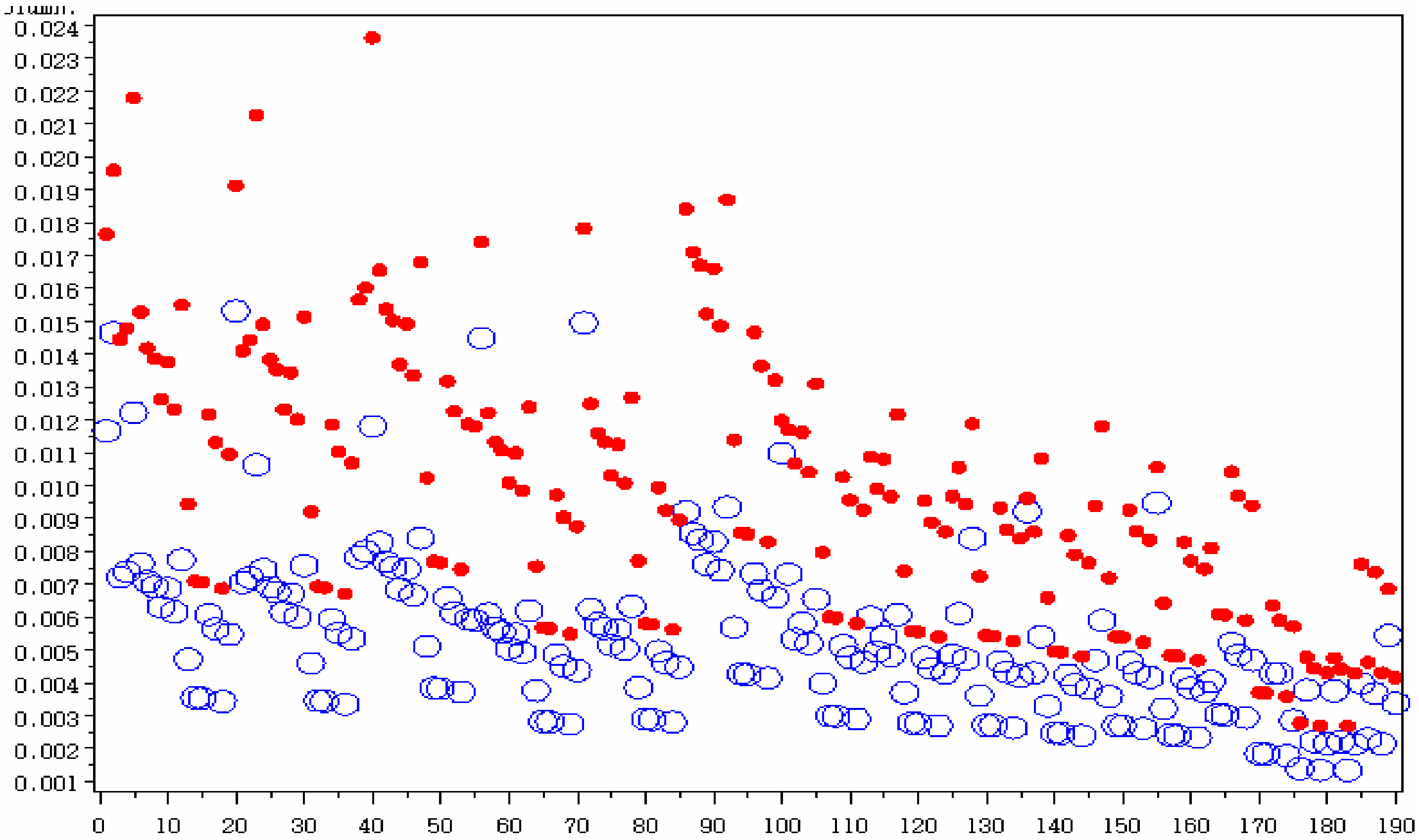
(Brewer’s method)



(OP I) $c = 0.5$


0.024
0.023
0.022
0.021
0.020
0.019
0.018
0.017
0.016
0.015
0.014
0.013
0.012
0.011
0.010
0.009
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190

(OP II) $c = 0.5$


0.024
0.023
0.022
0.021
0.020
0.019
0.018
0.017
0.016
0.015
0.014
0.013
0.012
0.011
0.010
0.009
0.008
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190

Concluding Remarks

- ✓ We have used the algorithm of Harvey (1976) for estimating the superpopulation model and examined the capacity of OWSP to yield a small variance.

- ✓ It seems that OWSP developed by Kim, Heeringa, and Solenberger (2006) is preferable to TDDP in terms of the efficiency. We also observe that OWSP shows better results as the population size increases.

- ✓ Since the objective function in optimization problem has a simple linear form, finding a solution, the sample selection probabilities, is not complicated.

- ✓ It appears that the linear constraints involving the value of c are quite useful to reduce the variance. But it requires a careful choice of the value.

- ✓ The empirical studies for more natural populations may be useful.

- ✓ It may need to study the stability of the variance estimator as well as the efficiency.

- ✓ The nonlinear superpopulation model might be adopted to develop a new sample selection procedure.

Contact Information

Sun-Woong Kim
Sunwk@dongguk.edu

Steven G. Heeringa
sheering@isr.umich.edu

Sung-Joon Hong
hsj8129@dongguk.edu

Peter W. Solenberger
pws@isr.umich.edu