J. Ind. Soc. Agril. Statist.
60(2), 2006 : 108-117

Estimation of Finite Population Mean using Ranked Set Two Stage Sampling Designs

U.C. Sud and Dwijesh Chandra Mishra
Indian Agricultural Statistics Research Institute, New Delhi
(Received : December, 2005)

SUMMARY

The Ranked Set Sampling (RSS) procedure for the estimation of finite population mean has been extended to the case of two stage sampling design. Three different cases have been considered : (a) Simple Random Sampling (SRS) at first stage and RSS at second stage, (b) RSS at first stage and SRS at second stage and (c) RSS at both the stages of sampling. It is empirically demonstrated that the use of RSS results in gain in precision of estimate of finite population mean over the SRS procedure.

Key words : Ranked set sampling, Two stage sampling, Finite population sampling.

1. INTRODUCTION

In statistical settings where actual measurements of the sample observations are difficult or costly or time consuming or destructive etc. but acquisition and subsequent ranking of the potential sample data is relatively easy, improved methods of statistical inference can result from using Ranked Set Sampling (RSS) technique.

The method of RSS was first introduced by McIntyre (1952) to improve upon simple random sampling for situations where some preliminary ranking of sampled units is possible. The basic idea was to randomly partition the sampled units into small groups and each member of the individual group is ranked relative to other members of the group. Based on this ranking one member from each group is selected for quantification.

The method involves selecting 'm' samples, each of size m, and ordering each of the sample by eye or by some other relatively inexpensive means, without actual measurement of the individuals. Then the smallest ordered observation from the 1st sample is accurately measured, as is the 2nd smallest observation from the 2nd. The process is continued until the largest observation from the mth sample is measured. This constitute one cycle.

The entire cycle is repeated n times, (because accurate judgment ordering of a large number of observations would be difficult in most experimental situations, therefore an increase in sample size is typically implemented by increasing n, the number of replication of cycle, rather than m) until altogether N = mn observations have been quantified.

Let $X_{11}, X_{12}, \ldots, X_{1m}, X_{22}, \ldots, X_{2m}, \ldots, X_{m1}, X_{m2}, \ldots, X_{mm}$ be independent random variables all having the same cumulative distribution function F(x). Also let $X_{i(1)}, X_{i(2)}, \ldots X_{i(i)}, \ldots, X_{m(m)}$ be the ranked set sample obtained (considering one cycle only) as per the procedure described in the preceding paragraph. We suppose that there are no ranking errors. Then we select the units for 'n' number of cycles.

RSS works by creating an "artificially" stratified samples. RSS provides a more precise estimator of population mean than SRS. This is due to the fact that RSS results in a sample in which units are more evenly spaced. Since the units in RSS are more evenly spaced than SRS, the variance of RSS estimates is expected to be less than SRS estimates. It is also cost efficient in a given situation.

Most of the work related to RSS is based on a framework wherein the population under study is assumed to follow some specific distribution. Patil *et al.*

(1995) and Krishna (2002) extended the theory of RSS to the case of sampling from a finite population. However, the contributions made by Patil *et al.* (1995) and Krishna (2001) were limited to uni-stage sampling design. In this paper an attempt has been made to extend the procedure of RSS to two-stage sampling designs.

Three different cases have been studied. In the first case SRS is used at the 1st stage of sampling and RSS at the 2nd stage of sampling. Similarly, RSS is used at 1st stage and SRS at the 2nd stage in second case. In the third case RSS is used in both the stages of sampling. In each of the cases efficiency comparisons of RSS based estimators have been made with SRS based estimators when the sampling is SRS at both the stages of sampling with the help of real data.

2. RSS FOR TWO-STAGE SAMPLING DESIGN

Let there be a finite population of N primary stage units, a-th primary stage unit is of size M. Let x_{ab} be the value of unit pertaining to b-th secondary stage unit (ssu) of a-th primary stage unit (psu).

Denote by

$$\overline{X}_a = \frac{1}{M}\sum_{b=1}^{M} x_{ab}$$

= mean per ssu in the a-th psu

$$\overline{X} = \frac{1}{NM}\sum_{a=1}^{N}\sum_{b=1}^{M} x_{ab}$$

= Population mean

Case 1: SRS at first stage and RSS at second stage

Let a sample of size 'n' be drawn from 'N' by SRSWOR. Also, let a set of size m be selected at random and without replacement from M using RSS.

Without any loss of generality we assume that

$$(x_{a1} \le x_{a2} \le \ldots \le x_{aM}) \ \forall \ a = 1,2,\ldots,N$$

Define the event

$$\{k \Rightarrow s\}$$

such that the k-th ranked unit in the subset is the s-th ranked unit in the population of ssu.

Also write

$$A_{ak}^{s} = \Pr\{k \Rightarrow s\}$$

and let A_{ak} denotes the M-dimensional column vector having A_{ak}^{s} as its s-th component

$$A'_{ak} = [A_{ak}^{1}\ A_{ak}^{2}\ .\ \ .\ \ .\ A_{ak}^{s}\ .\ \ .\ \ .\ A_{ak}^{N}]$$

It may be noted that A_{ak}^{s} is given by

$$A_{ak}^{s} = \frac{\binom{s-1}{k-1}\binom{M-s}{m-k}}{\binom{M}{m}}, \quad s = 1,2,\ldots,M$$

If $x_{a(k:m)}$ is the quantification of the k-th ranked unit from the set, then

$$\begin{aligned} E[x_{a(k:m)}] &= \overline{X}_{a(k:m)} \\ &= \sum_{s=1}^{M} x_{as} \Pr[x_{a(k:m)} = x_{as}] \\ &= \sum_{s=1}^{M} x_{as} \Pr(\{k \Rightarrow s\}) \\ &= \sum_{s=1}^{M} x_{as} A_{ak}^{s} \\ &= A'_{ak} x_a \end{aligned}$$

$$\begin{aligned} V[x_{a(k:m)}] &= \sigma^2_{xa(k:m)} \\ &= A'_{ak} x_a^2 - (A'_{ak} x_a)^2 \end{aligned}$$

where, x_a^2 is the component wise square of x_a.

Next, we study the joint distribution of the order statistics from two disjoint sets. Let two disjoint sets each of size m be drawn without replacement from M.

Write

$$\{k \Rightarrow s, j \Rightarrow t\}$$

for the event that the k-th ranked unit from set 1 has rank s and the j-th ranked unit from set 2 has rank t in the population of size M.

We define

$$B_{akj}^{st} = \Pr\{k \Rightarrow s, j \Rightarrow t\}$$

Following Patil *et al.* (1994), it may be seen that

$$B^{st}_{akj} = \sum_{\lambda=0}^{m-k} \frac{\binom{s-1}{k-1}\binom{t-s-1}{\lambda}\binom{M-t}{m-k-\lambda}\binom{t-1-k-\lambda}{j-1}\binom{M-t-m+k+\lambda}{m-j}}{\binom{M}{m,m}}$$

Let B_{ajk} be the $M \times M$ matrix with B^{st}_{akj} as its (s,t)-th component. Notice that $B_{akj} = B'_{ajk}$, since $B^{st}_{akj} = B^{ts}_{ajk}$.

Let $x_{a(k:m)1}$ and $x_{a(j:m)2}$ be the quantification of the k-th and j-th ranked units from set 1 and set 2, respectively. Then

$$\begin{aligned} E[x_{a(k:m)1}, x_{a(j:m)2}] &= \sum_{s,t=1}^{M} x_{as} x_{at} \Pr[x_{a(k:m)1} = x_{as}, x_{a(j:m)2} = x_{at}] \\ &= \sum_{s,t=1}^{M} x_{as} x_{at} B^{st}_{akj} \\ &= x'_a B_{akj} x_a \end{aligned}$$

Thus, the covariance between $x_{a(k:m)1}, x_{a(j:m)2}$ is given by

$$\begin{aligned} \mathrm{Cov}[x_{a(k:m)1}, x_{a(j:m)2}] &= C_{akj} \\ &= x'_a (B_{akj} - A_{ak} A'_{aj}) x_a \end{aligned}$$

Let mr sets, each of size m, be selected randomly using RSS and without replacement from the a-th psu. Let the lowest ranked unit be quantified in each of the first 'r' sets as

$$X_{a(1:m)1}, X_{a(1:m)2}, \ldots, X_{a(1:m)r}$$

In each of the next r sets, the second ranked unit is quantified to give

$$X_{a(2:m)1}, X_{a(2:m)2}, \ldots, X_{a(2:m)r}$$

This process continues until the highest ranked unit is quantified in each of the last r sets:

$$X_{a(m:m)1}, X_{a(m:m)2}, \ldots, X_{a(m:m)r}$$

Theorem 1. The estimator

$$\hat{\bar{X}}_{RSS1} = \frac{1}{nrm} \sum_{a=1}^{n} \sum_{o=1}^{r} \sum_{k=1}^{m} x_{a(k:m)o} \qquad (2.1)$$

is unbiased and variance of $\hat{\bar{X}}_{RSS1}$ is given by

$$\mathrm{Var}(\hat{\bar{X}}_{RSS1}) = \left(\frac{1}{n} - \frac{1}{N}\right) S^2_{xa} + \frac{1}{nrmN} \sum_{a=1}^{N} [\{\sigma^2_{xa} \frac{(M-1-mr)}{(M-1)} - \bar{\gamma}_a\}] \qquad (2.2)$$

where

$$\begin{aligned} S^2_{xa} &= \frac{1}{(N-1)} \sum_{a=1}^{N} (\bar{X}_a - \bar{X})^2 \\ \bar{\gamma}_a &= \frac{m!(m-1)!}{M(M-1)\ldots(M-2m+1)} \gamma_a \\ \gamma_a &= (x_a - \mu_a)' \Gamma (x_a - \mu_a) \end{aligned}$$

The matrix Γ is symmetric with zeroes on the diagonal, it is calculated by

$$\Gamma = \binom{M}{m,m} \sum_{k=1}^{m} B^{*}_{akk}$$

Proof :

To prove that the estimator $\hat{\bar{X}}_{RSS1}$ is unbiased, we proceed as follows

$$E_1 E_2(\hat{\bar{X}}_{RSS1}) = E_1 E_2[\frac{1}{nrm} \sum_{a=1}^{n} \sum_{o=1}^{r} \sum_{k=1}^{m} x_{a(k:m)o}]$$

where E_2 is the expectation with respect to RSS in a given primary stage unit and E_1 is the unconditional expectation at the first stage of sampling.

$$\begin{aligned} &= E_1 \frac{1}{n} \sum_{a=1}^{n} E_2[\frac{1}{rm} \sum_{o=1}^{r} \sum_{k=1}^{m} x_{a(k:m)o}] \\ &= E_1 \frac{1}{n} \sum_{a=1}^{n} \frac{1}{rm} \sum_{o=1}^{r} \sum_{k=1}^{m} E_2[x_{a(k:m)o}] \\ &= E_1 \frac{1}{n} \sum_{a=1}^{n} \frac{1}{rm} \sum_{o=1}^{r} \sum_{k=1}^{m} A'_{ak} x_a \\ &= E_1 \frac{1}{n} \sum_{a=1}^{n} \frac{1}{rm} \sum_{o=1}^{r} \sum_{k=1}^{m} \sum_{s=1}^{M} A^{s}_{ak} x_{as} \\ &= E_1 \frac{1}{n} \sum_{a=1}^{n} \frac{1}{m} \sum_{s=1}^{M} (\sum_{k=1}^{m} A^{s}_{ak}) x_{as} \end{aligned}$$

* A computer program has been made in language turbo 'c' to calculate Γ.

$$= E_1 \frac{1}{n}\sum_{a=1}^{n}\frac{1}{M}\sum_{s=1}^{M} x_{as}$$

$$= E_1 \frac{1}{n}\sum_{a=1}^{n}\overline{X}_a$$

$$= \frac{1}{N}\sum_{a=1}^{N}\overline{X}_a$$

$$= \overline{X}$$

Implying thereby that $\hat{\overline{X}}_{RSS1}$ is unbiased.

The variance of $\hat{\overline{X}}_{RSS1}$ can be obtained as follows

$$V(\hat{\overline{X}}_{RSS1}) = E_1V_2(\hat{\overline{X}}_{RSS1}) + V_1E_2(\hat{\overline{X}}_{RSS1})$$

where V_1 and V_2 can be defined on similar lines as E_1 and E_2.

Following Patil *et al.* (1994)

$$V_2(\hat{\overline{X}}_{RSS1})$$

$$= Var\left[\frac{1}{nrm}\sum_{a=1}^{n}\sum_{o=1}^{r}\sum_{k=1}^{m} x_{a(k:m)o}\right]$$

$$= \frac{1}{n^2}\sum_{a=1}^{n}[\frac{1}{r^2m^2}\{r\sum_{k=1}^{m}\sigma^2_{a(k:m)} + r^2\sum_{k=1}^{m}\sum_{j=1}^{m}C_{akj} - \sum_{k}^{m}C_{akk}\}]$$

$$= \frac{1}{n^2}\sum_{a=1}^{n}[\frac{1}{rm^2}\{m\sigma^2_{x_a}\frac{(M-1-mr)}{(M-1)}$$

$$-\sum_{k=1}^{m}(\overline{X}_{a(k:m)} - \overline{X}_a)^2 - \sum_{k=1}^{m}C_{akk}\}]$$

It may be noted that the last term must remain unchanged if the population is centered, i.e., if x_{as} is replaced by $x_{as} - \overline{X}_a$ and $\overline{X}_{a(k:m)}$ is replaced by $\overline{X}_{a(k:m)} - \overline{X}_a$. Since

$$C_{akk} = x'_a(B_{akk} - A_{ak}A'_{ak})x_a$$

$$= x'_aB_{akk}x_a - \mu^2_{a(k:m)}$$

which becomes after centering

$$C_{akk} = (x_a - \overline{X}_a)'B_{akk}(x_a - \overline{X}_a) - (\overline{X}_{a(k:m)} - \overline{X}_a)^2$$

Thus

$$V_2(\hat{\overline{X}}_{RSS1}) = \frac{1}{n^2}\sum_{a=1}^{n}[\frac{1}{rm}\{\sigma^2_{x_a}\frac{(M-1-mr)}{(M-1)}$$

$$-\frac{1}{m}\sum_{k=1}^{m}(x_a - \overline{X}_a)'B_{akk}(x_a - \overline{X}_a)\}]$$

$$= \frac{1}{n^2}\sum_{a=1}^{n}[\frac{1}{rm}\{\sigma^2_{x_a}\frac{(M-1-mr)}{(M-1)} - \overline{\gamma}_a\}]$$

Now

$$E_1V_2(\hat{\overline{X}}_{RSS1})$$

$$= E_1[\frac{1}{n^2}\sum_{a=1}^{n}[\frac{1}{rm}\{\sigma^2_{x_a}\frac{(M-1-mr)}{(M-1)} - \overline{\gamma}_a\}]$$

$$= \frac{1}{nN}\sum_{a=1}^{N}[\frac{1}{rm}\{\sigma^2_{x_a}\frac{(M-1-mr)}{(M-1)} - \overline{\gamma}_a\}] \qquad (2.3)$$

Also

$$V_1E_2(\hat{\overline{X}}_{RSS1})$$

$$= V_1[\frac{1}{n}\sum_{a=1}^{n}E_2\{\frac{1}{rm}\sum_{o=1}^{r}\sum_{k=1}^{m}x_{a(k:m)o}\}]$$

$$= V_1\{\frac{1}{n}\sum_{a=1}^{n}\overline{X}_a\}$$

$$= \left(\frac{1}{n} - \frac{1}{N}\right)S^2_{x_a} \qquad (2.4)$$

Combining (2.3) and (2.4), we obtain the desired result.

Hence the proof.

Case2: RSS at first stage and SRS at second stage

Assume that a sample of size 'm' is selected by SRSWOR from the a-th psu a = 1, 2, ..., N. Further, we assume that a set of size 'n' is selected from 'N' by RSS. Also, as in Case 1, we assume that the psu's are increasingly arranged.

Define the event

$$\{a \Rightarrow s\}$$

such that the a-th ranked unit in the subset is the s-th ranked unit in the population of psu's.

Define $A^s_a = Pr\{a \Rightarrow s\}$

$$A'_a = [A_a^1 \ A_a^2 \ . \ . \ . \ A_a^s \ . \ . \ . \ A_a^N]$$

be the $N \times 1$ row vector having A_a^s as its s-th component.

It may be noted that A_a^s is given by

$$A_a^s = \frac{\binom{s-1}{a-1}\binom{N-s}{n-a}}{\binom{N}{n}}, \ s=1,2,\ldots,N; \ a=1,2,\ldots,n$$

We denote by

$$\bar{x}_{(a:n)} = \frac{1}{m}\sum_{b=1}^{m} x_{(a:n)b} = \text{sample mean for the a-th psu}$$

$$E_2[\bar{x}_{(a:n)b}] = \frac{1}{M}\sum_{b=1}^{M} x_{(a:n)b}$$

$$= \overline{X}_{s(a:n)}$$

$$E_1[\overline{X}_{(a:n)}] = \overline{X}_{(a:n)}$$

$$= \sum_{s=1}^{N} \overline{X}_s \Pr\{X_{(a:n)} = x_s\}$$

$$= \sum_{s=1}^{N} \overline{X}_s \Pr\{a \Rightarrow s\}$$

$$= \sum_{s=1}^{N} \overline{X}_s A_a^s$$

$$= A'_a \overline{X}$$

$$V_1(\overline{X}_{(a:n)}) = \sigma^2_{x(a:n)}$$

$$= A'_a \overline{X}^2 - (A'_a \overline{X})^2$$

where $\overline{X}^2$ is the component wise square of $\overline{X}$.

To study the joint distribution of the order statistics from disjoint sets each of size 'n' drawn by without replacement using RSS, let

$$\{a \Rightarrow s, c \Rightarrow t\}$$

be the event that the a-th ranked unit from set 1 has rank s in the population and the c-th ranked unit from set 2 has rank t in the population.

Define

$$B_{ac}^{st} = \Pr\{a \Rightarrow s, c \Rightarrow t\}$$

Then again following, Patil *et al.* (1995)

$$B_{ac}^{st} = \sum_{\lambda=0}^{n-a} \frac{\binom{s-1}{a-1}\binom{t-s-1}{\lambda}\binom{M-t}{m-a-\lambda}\binom{t-1-a-\lambda}{c-1}\binom{M-t-m+a+\lambda}{m-c}}{\binom{N}{n,n}}$$

Let B_{ac} be the $N \times N$ matrix with B_{ac}^{st} as its (s,t)-th component. Notice that $B_{ac} = B'_{ca}$ since

$$B_{ac}^{st} = B_{ca}^{ts}$$

Let $\bar{x}_{(a:n)1}$ and $\bar{x}_{(c:n)2}$ be the quantification of the a-th and c-th ranked units from set 1 and set 2, respectively. Then

$$E[\bar{x}_{(a:n)1}, \bar{x}_{(c:n)2}]$$

$$= \sum_{s,t=1}^{N} \bar{x}_s \bar{x}_t \Pr[\bar{x}_{(a:n)} = \bar{x}_s, \bar{x}_{(c:n)} = \bar{x}_t]$$

$$= \sum_{s,t=1}^{N} \bar{x}_s \bar{x}_t B_{ac}^{st}$$

$$= \bar{x}' B_{ac} \bar{x}$$

Covariance between $\bar{x}_{(a:n)}, \bar{x}_{(c:n)}$ is given by

$$\text{Cov}[\bar{x}_{(a:n)}, \bar{x}_{(c:n)}] = C_{ac}$$

$$= \bar{x}'(B_{ac} - A_a A'_c)\bar{x}$$

Moments of the estimator of population mean

Let nr sets each of size n be selected randomly and without replacement from a population of N psu's. Let the lowest ranked unit be quantified in each of the first r sets

$$\overline{X}_{(1:n)1}, \overline{X}_{(1:n)2}, \ldots, \overline{X}_{(1:n)r}$$

Similarly, in each of the next r sets, the second ranked unit is quantified to give

$$\overline{X}_{(2:n)1}, \overline{X}_{(2:n)2}, \ldots, \overline{X}_{(2:n)r}$$

This process continues until the highest ranked unit is quantified in each of the last r sets

$$\overline{X}_{(m:n)1}, \overline{X}_{(m:n)2}, \ldots, \overline{X}_{(m:n)r}$$

Thus, the proposed estimator of population mean, when the sample at the first stage is selected by RSS and at the second stage by SRS, is given by

$$\hat{\overline{X}}_{RSS2} = \frac{1}{nrm}\sum_{a=1}^{n}\sum_{o=1}^{r}\sum_{b=1}^{m} x_{((a:n)o,b)} \qquad (2.5)$$

On the same lines as in Case 1, it can be shown that $\hat{\overline{X}}_{RSS2}$ is unbiased, the variance of $\hat{\overline{X}}_{RSS2}$ is given by

$$V(\hat{\overline{X}}_{RSS2}) = \frac{1}{rn^2}\{\frac{N-1-nr}{N-1}\sigma_x^2 - \overline{\gamma}\} + \frac{1}{rnN}\sum_{a=1}^{N}\{(\frac{1}{m}-\frac{1}{M})S^2_{x(a:n)}\} \qquad (2.6)$$

where

$$S^2_{x(a:n)} = \frac{1}{(M-1)}\sum_{b=1}^{M}(x_{(a:n)b} - \overline{X}_{(a:n)})^2$$

Case 3 : RSS at both the stages

We assume that both the psu's and the ssu's within a psu are increasingly arranged.

We may write

$$\overline{X} = (\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_N)$$

where $\overline{X}_i$ is the mean per ssu in a psu.

Let a set of size m be selected at random and without replacement from the u-th psu (size of u-th psu being M) as already stated.

Define the event $\{i \Rightarrow s\}$ to mean that the i-th ranked unit in the subset of size m is the s-th ranked unit in the population. Also, write

$$A_i^s = \Pr\{i \Rightarrow s\}$$

It may be seen that

$$A_i^s = \frac{\binom{s-1}{i-1}\binom{M-s}{m-i}}{\binom{M}{m}}$$

Let A_i denote the M dimensional column vector having A_i^s as its s-th component.

Similarly, let a set of size 'n' be selected at random and without replacement, also by RSS, from 'N'.

Define the event

$$\{k \Rightarrow u\}$$

to mean that the k-th ranked unit in the subset is the u-th ranked unit in the population. Also, write

$$D_k^u = \Pr\{k \Rightarrow u\}$$

where

$$D_k^u = \frac{\binom{u-1}{k-1}\binom{N-u}{n-k}}{\binom{N}{n}}, \quad u=1,2,\ldots,N;\ k=1,2,\ldots,n$$

Let D_k denotes the N dimensional column vector having D_k^u as its u-th component.

Denote by $X_{(k:n,\, i:m)}$ the quantification of the i-th ranked second stage unit from the u-th psu. Let E_1 and E_2 denote the expectation at the first stage and second stage of sampling respectively, then

$$E_2[X_{(k:n,i:m)}] = \overline{X}_{(k:n,i:m)}$$

$$= \sum_{s=1}^{M} x_{(k:n,s)} \Pr\{X_{(k:n,i:m)} = x_{k:n,s}\}$$

$$= \sum_{s=1}^{M} x_{(k:n,s)} \Pr\{i \Rightarrow s/u\}$$

$$= \sum_{s=1}^{M} x_{(k:n,s)} A_{ui}^s$$

$$= A'_{ui} x_{k:n}$$

where $x_{k:n} = (x_{k:n,1}, x_{k:n,2}, \ldots, x_{k:n,m})$

On the same lines

$$V_2(X_{k:n,i:m}) = \sigma^2_{xuk:n,i:m)}$$

$$= A'_{ui} x^2_{k:n} - \{A'_{ui} x_{k:n}\}^2$$

where $x^2_{k:n}$ is the component-wise square of $x_{k:n}$.

To study the joint distribution of order statistics from the disjoint sets within a psu, we proceed as follows

Let two disjoint sets of size m each be drawn without replacement from the u-th psu's.

We write

$$\{i \Rightarrow s, j \Rightarrow t\}$$

for the event that the i-th ranked unit from a set of size m has the s-th rank and j-th ranked unit has t-th rank in the population.

Define

$$B_{uij}^{st} = \Pr\{i \Rightarrow s, j \Rightarrow t\}$$

If s < t, then

$$B_{uij}^{st} = \sum_{\lambda=0}^{m-i} \frac{\binom{s-1}{i-1}\binom{t-s-1}{\lambda}\binom{M-t}{m-i-\lambda}\binom{t-1-i-\lambda}{j-1}\binom{M-t-m+i+\lambda}{m-j}}{\binom{M}{m,m}}$$

Let B_{ij} be the $M \times M$ matrix with B_{uij}^{st} as its s-th component.

Notice that $B_{ij} = B'_{ji}$ since $B_{uij}^{st} = B_{uji}^{st}$

To study the joint distribution of order statistics from disjoint sets at the psu level, we proceed as follows

Let two disjoint sets of size 'n' each be drawn without replacement from the population of 'N' psu's.

We write

$$\{k \Rightarrow s', l \Rightarrow t'\}$$

for the event that the k-th ranked unit from a set of size n has s'-th ranked and l-th ranked unit has t'-th rank in the population.

Define

$$D_{kl}^{s't'} = \Pr\{k \Rightarrow s', l \Rightarrow t'\}$$

If $s' < t'$ then

$$D_{kl}^{s't'} = \sum_{\lambda=0}^{n-k} \frac{\binom{s'-1}{k-1}\binom{t'-s'-1}{\lambda}\binom{N-t'}{n-k-\lambda}\binom{t'-1-k-\lambda}{l-1}\binom{N-t'-n+k+\lambda}{n-l}}{\binom{N}{n,n}}$$

Let $X_{(k:n,i:m)1}, X_{(k:n,j:m)2}$ be the quantification of the i-th and the j-th ranked units from the set 1 and set 2 respectively of the u-th psu.

Then

$$E_2[X_{(k:n,i:m)1}, X_{(k:n,j:m)2}]$$

$$= \sum_{s,t=1}^{M} x_{(k=u,s)} x_{(k=u,t)} P\{X_{(k:n,i:m)} = x_{us}, X_{(k:n,i:m)} = x_{ut}\}$$

$$= \sum_{s,t=1}^{M} x_{(k=u,s)} x_{(k=u,t)} B_{uij}^{st}$$

$$= x'_{k:n} B_{uij} x_{k:n}$$

Thus, Covariance between elements from two disjoint sets is given by

$$\mathrm{Cov}[X_{(k:n,i:m)1}, X_{(k:n,i:m)2}]$$

$$= x'_{k:n}(B_{uij} - A_{ui}A'_{uj})x_{k:n}$$

Let $\overline{X}_{(k:n)1}, \overline{X}_{(k:n)2}$ be the quantification of the k-th and l-th ranked units from set 1 and set 2 respectively of the two psu's.

Then

$$E_1(\overline{X}_{(k:n)1}, \overline{X}_{(k:n)2})$$

$$= \sum_{s',t'}^{N} \overline{x}_k \overline{x}_l \Pr\{\overline{X}_{(k:n)1} = \overline{x}_{s'}, \overline{X}_{(l:n)2} = \overline{x}_{t'}\}$$

$$= \sum \overline{x}_{s'} \overline{x}_{t'} D_{kl}^{s't'}$$

$$= \overline{x}' D_{kl} \overline{x}$$

$$\mathrm{Cov}[\overline{x}_{(k:n)1}, \overline{x}_{(l:n)2}]$$

$$= C_{kl} = \overline{x}'(D_{kl} - D_k D'_l)\overline{x}$$

Let r_1 sets each of size n be selected randomly and without replacement from N. Let the lowest ranked unit be quantified in each of the first r_1 sets.

Similarly in each of the next r_1 sets the second ranked unit is quantified.

Likewise k-th ranked unit is quantified in each of the last r_1 sets.

As in the case of primary stage units a sample of mr_2 second stage units is selected from the q-th ranked quantified unit. These mr_2 units are quantified exactly the same way as in the case of primary stage units. The units so quantified can be represented as

$$X_{[(1:n)1,(1:m)1]}, X_{[(1:n)1,(1:m)r_2]} \ldots X_{[(1:n)1,(1:m)r_2]}$$

$$X_{[(1:n)1,(2:m)1]}, X_{[(1:n)1,(2:m)2]}, \ldots, X_{[(1:n)1,(2:m)r_2]}$$

$$\vdots$$

$$X_{[(1:n)1,(m:m)1]}, X_{[(1:n)1,(m:m)2]}, \ldots, X_{[(q:n)1,(m:m)r_2]}$$

$$\vdots$$

$$X_{[(n:n)r_1,(m:m)1]}, X_{[(n:n)r_1,(m:m)2]}, \ldots, X_{[(n:n)r_1,(m:m)r_2]}$$

The ranked set estimator of $\overline{X}$ is the average of these quantifications.

$$\hat{\overline{X}}_{RSS3} = \frac{1}{r_1 n}\sum_{o=1}^{r_1}\sum_{k=1}^{n}\frac{1}{r_2 m}\sum_{v=1}^{r_2}\sum_{i=1}^{m} X_{[(k:n)o;(i:m)v]} \tag{2.7}$$

This is unbiased and its variance is given by

$$V(\hat{\overline{X}}_{RSS3}) = \frac{1}{n^2 r_1}\left[\frac{N-1-nr_1}{N-1}\sigma_x^2 - \overline{\gamma}\right] + \frac{1}{Nnr_1r_2}\sum_{k=1}^{N}\frac{1}{m^2}\left[\frac{M-1-r_2m}{M-1}\sigma_{xk}^2 - \overline{\gamma}_k)\right] \tag{2.8}$$

3. EMPIRICAL STUDY

For the purpose of comparing the RSS and SRS based estimators, an empirical study was carried out where in a part of the data of wheat crop for an experimental station as given in Singh *et al.* (1979) was taken. The data comprised of 9 fields, each field having 4 plots. (Set I) For RSS protocol, plots in each field were ranked according to the perceived weight of wheat yield. Using this data, estimators of population mean based on RSS and SRS were considered for the three cases dealt with earlier.

Another data set given in Singh and Mangat (1996) on outstanding loans of farmers affiliated to cooperatives was utilized to compare the performance of RSS and SRS based estimators. (Set II)

Finally data on number of persons in a household given in Raj (1971) was also utilized to compare the performance of RSS and SRS based estimators. (Set III)

Table 2.1. Percent gain in precision of RSS based estimators over SRS based estimators

Case	Stage	Design	Estimator	S.E. of the estimator	Percent gain in precision
Set I					
1	1	SRS	$\hat{\overline{X}}_{RSS1}$	5.39	10.21
	2	RSS			
2	1	RSS	$\hat{\overline{X}}_{RSS2}$	5.60	1.85
	2	SRS			
3	1	RSS	$\hat{\overline{X}}_{RSS3}$	5.33	12.46
	2	RSS			
4	1	SRS	$\hat{\overline{X}}_{SRS}$	5.94	—
Set II					
1	1	SRS	$\hat{\overline{X}}_{RSS1}$	1.03	10.21
	2	RSS			
2	1	RSS	$\hat{\overline{X}}_{RSS2}$	1.11	1.85
	2	SRS			
3	1	RSS	$\hat{\overline{X}}_{RSS3}$	1.01	12.46
	2	RSS			
4	1	SRS	$\hat{\overline{X}}_{SRS}$	1.14	—
Set III					
1	1	SRS	$\hat{\overline{X}}_{RSS1}$	0.199	10.21
	2	RSS			
2	1	RSS	$\hat{\overline{X}}_{RSS2}$	0.205	1.85
	2	SRS			
3	1	RSS	$\hat{\overline{X}}_{RSS3}$	0.194	12.46
	2	RSS			
4	1	SRS	$\hat{\overline{X}}_{SRS}$	0.230	—

It may be seen from Table 2.1 that the RSS based estimators, whether RSS used at first stage or at second stage or at both the stages of sampling, scores over the SRS based estimators. The gain in precision for all the three data sets is maximum when RSS is used at both the stages of sampling. Thus, the results of the study clearly demonstrate that the RSS procedure is more efficient than SRS in case of two stage sampling design also.

REFERENCES

Al-Saleh, M.F. and Al-Omari, A.I. (2002). Multistage ranked set sampling. *J. Statist. Plann. Inf.*, **102**, 273- 286.

Al-Saleh, M.F. and Al-Shrafat, K. (2001). Estimation of average milk yield using ranked set sampling. *Environmetrics*, **12**, 395-399.

Chuiv, N.N. and Sinha, B.K. (1998). *On Some Aspects of Ranked Set Sampling in Parametric Estimation.* Handbook of Statistics, **17**, 337-377, N. Balakrishna and Rao, C.R. Eds., North Holland, New York.

Chuiv N., Sinha, B.K., and Wu, Z. (1994). *Estimation of the location parameters of a cauchy distribution using a ranked set sample.* Technical Report, University of Maryland, Baltimore County.

Krishna, Pravin (2002). *Some aspects of ranked set sampling from finite population.* M.Sc.Thesis of I.A.R.I., New Delhi-12.

Patil, G.P., Sinha, A.K. and Taillie, C. (1993). Ranked set sampling from a finite population in the presence of a trend on a site. *J. Appl. Statist. Sci.*, **1(1)** 51-65.

Patil, G.P., Sinha, A.K. and Taillie, C. (1994). *Ranked Set Sampling.* Handbook of Statistics, **12**, (eds. Patil, G.P. and Rao, C.R.), 167-198, North-Holland, Amsterdam.

Patil, G.P., Sinha, A.K. and Taillie, C. (1995). Finite population corrections for ranked set sampling. *Ann. Math. Statist.*, **47(4)**, 621-636.

Raj, D. (1971). *The Design of Sample Surveys.* Mcgraw-Hill Book Co., New York.

Singh, D., Singh, P. and Kumar, P. (1979). *Hand Book on Sampling Methods.* Indian Agricultural Statistics Research Institute, New Delhi.

Singh, R. and Mangat, N.P.S. (1996). *Elements of Survey Sampling.* Kluwer Academic Publisher, 388.

Sukhatme, P.V. and Sukhatme, B.V. (1970). *Sampling Theory of Surveys with Application.* American Iowa State University Press.

APPENDIX

A computer program to calculate T matrix

```
#include<stdio.h>
#include<conio.h>
#include<math.h>
fact(int,int);
void main()
{
FILE *fp;
fp = fopen("c:/output.doc","w");
float aa[6][30][30];
int n,m,i,l,ii,jj;
int a,b,c,d,e,f,g,h,o,p,q1,r1,q2,r2;
float u,v,w,x,y,z1,z2,cc,c1;
float j,k;
clrscr( );
scanf("%d%d",&n,&m);
for(i=1;i<=m;i++){
for(ii=1;ii<=n;ii++){
for(jj=1;jj<=n;jj++){
k=0.0;
for(l=0;l<=(m-i);l++)
{
a=ii-1;
b=i-1;
c=jj-ii-1;
d=l;
e=n-jj;
f=m-i-l;
g=jj-1-i-l;
h=i-1;
o=n-jj-m+i+l;
p=m-i;
q1=n;
r1=m;
q2=n-m;
r2=m;
u=fact(a,b);
v=fact(c,d);
w=fact(e,f);
x=fact(g,h);
y=fact(o,p);
z1=fact(q1,r1);
z2=fact(q2,r2);
j=float((u*v*w*x*y)/(z1*z2));
```

```
k=k+j;
}

aa[i][ii][jj]=k;
}}}
for(i=1;i<=m;i++){
fprintf(fp,"\n");
for(ii=1;ii<=n;ii++){
fprintf(fp,"\n");
for(jj=1;jj<=n;jj++){
fprintf(fp,"%.2f\t",aa[i][ii][jj]);
}}}
fprintf(fp,"\n");
for(ii=1;ii<=n;ii++){
fprintf(fp,"\n");
for(jj=1;jj<=n;jj++){
cc=0.0;
for(i=1;i<=m;i++){
cc=cc+aa[i][ii][jj];
}
fprintf(fp,"%.2f\t",cc);
}}
getch( );
fclose(fp);
}
fact(int x,int y)
{
if(x==y){
return 1;
}
else if(x<y){
return 0;
}
else
{
int i,j,k;
float f;
long double c,d,e;
c=1;
for(i=1;i<=x;i++){
c=c*i;
}
d=1;
for(j=1;j<=y;j++){
d=d*j;
}
e=1;
for(k=1;k<=(x-y);k++){
e=e*k;
}
f=c/(d*e);
return(f);
}
}
```