Journal of Modern Applied Statistical Methods

Volume 3 | Issue 1 Article 12

5-1-2004

Stratified Extreme Ranked Set Sample With Application To Ratio Estimators

Hani M. Samawai
Sultan Qaboos University, hsamawi@squ.edu.om

Laith J. Saeid
Sultan Qaboos University

Journal of Modern Applied Statistical Methods
May 2004, Vol. 3, No. 1, 117-133

1538 – 9472/04/$95.00

Stratified Extreme Ranked Set Sample With Application To Ratio Estimators

Hani M. Samawi
Department of Mathematics & Statistics
Sultan Qaboos University

Laith J. Saeid
Department of Mathematics & Statistics
Sultan Qaboos University

Stratified extreme ranked set sample (SERSS) is introduced. The performance of the combined and separate ratio estimates using SERSS is investigated. Theoretical and simulation study are presented. Results indicate that using SERSS for estimating the ratios is more efficient than using stratified simple random sample (SSRS) and simple random sample (SRS). In some cases it is more efficient than ranked set sample (RSS) and stratified ranked set sample (SRSS), when the underlying distribution is symmetric. An application to real data on the bilirubin level in jaundice babies is introduced to illustrate the method.

Key words: Simple random sample; stratified random sample; ranked set sample; stratified ranked set sample; ratio estimation

Introduction

When sampling units in a study can be easily ranked compared to quantification. McIntyre (1952) proposed to use the mean of units based on a ranked set sample (RSS) to estimate the population mean. RSS is conducted by selecting random samples from the target population each of size r. Ranking each element within each set with respect to the random variable of interest. Then an actual measurement is taken of the element with the smallest rank from the first sample.

From the second sample an actual measurement is taken of the element with the second smallest rank, and the procedure is continued until the element with the largest rank is chosen for actual measurement from the r-th sample. Thus we obtain a total of r measured elements; one from each ordered sample of size r and this completed one cycle. The cycle may be repeated m times until $n = rm$ elements have been measured. These n elements form the ranked set sample data.

Hani M. Samawi is Associate Professor, Yarmouk University. Contact information is P.O. Box 36, Sultan Qaboos University, Al-khod, 123, Sultanate of Oman. Email: hsamawi@squ.edu.om.

Samawi et al. (1996) investigated variety of extreme ranked set samples (ERSS) for estimating the population means. Furthermore, Samawi (1996) introduced the principle of stratified ranked set sampling (SRSS); to improve the precision of estimating the population means in case of SSRS.

In many situations the quantity that is to be estimated from a random sample is the ratio of two variables both of which vary from unit to unit. For example, in a household survey, the average number of suits of clothes per adult male is the quantities of interest. Examples of this kind occur frequently when the sampling unit (the household) comprises a group or cluster of elements (adult males) and our interest is in the population mean per element.

Moreover, ratio appears in many other applications, for example, the ratio of loans for building purpose to total loans in a bank or the ratio of acres of wheat to total acres on a farm. Also, this method is to obtain increased precision of estimating the population mean or total by taking advantage of the correlation between an auxiliary variable X and the variable of interest Y.

In the literature, ratio estimators are used in case of SRS as well as in case of SSRS (for an example see Cochran, 1977). Also, SSRS is used in certain types of surveys because it combines the conceptual simplicity of simple random sample with potentially, significant gains in efficiency. It is a convenient technique to use whenever, one wish to ensure that a sample is representative of the population and also to obtain separate estimates for parameters of each sub-domain of the population. There are two methods for estimating ratios that are generally used when the sampling design is stratified random sampling, namely the combined ratio estimate and the separate ratio estimate. Moreover, Samawi and Muttlak (1996) used RSS to estimate the population ratio, and showed that it provided a more efficient estimator compared with using SRS.

Introduce in this article is the idea of stratified extreme ranked set sample (SERSS). Also, the use of the idea of SERSS is proposed to improve the precision of the two methods for estimating the ratio namely the combined ratio estimate and separate ratio estimate. Moreover, studied are the properties of these estimators and comparing them in different situations. Later in the article the principle of SERSS and its properties are introduced. Combined and separate ratio estimators using SERSS are then discussed followed by a simulation study and the results of the simulation including an illustration of the methods using real data about the bilirubin level in jaundice babies.

Methodology

Ranked set sample for bivariate elements

A modification of the above procedure used by Samawi and Muttlak (1996) for the estimation of the ratio. First choose r independent samples each of size r of independent bivariate elements from the target population. Rank each sample with respect to one of the variables *Y* or *X*. Suppose that the ranking is done on the variable *Y*. From the first sample an actual measurement is taken of the element with the smallest rank of *Y*, together with the value of the variable *X* associated with the smallest value of *Y*.

From the second sample an actual measurement is then taken of the element with the second smallest rank of *Y*, together with the value of the variable *X* associated with the second smallest value of *Y*. The procedure is continued until the element with the largest rank of *Y* is chosen for measurement from the *r-th* sample, together with the value of the variable *X* associated with the largest value of *Y*. The cycle may be repeated m times until $n = rm$ bivariate elements have been measured. Note that we assume that the ranking of the variable *Y* will be perfect, while the ranking of the variable *X* will be with errors in ranking, or at worst of a random order if the correlation between *Y* and *X* is close to zero.

Stratified ranked set sample

For the *h-th* stratum, first choose r_h independent samples each of size r_h of independent elements from the *h-th* subpopulations, $h = 1, 2, ..., L$. Rank each sample within each stratum, then use the same sampling scheme described above to obtain L independent RSS samples of sizes r_1, r_2, …, r_L respectively. Note that $r_1 + r_2 + \ldots + r_L = r$. This complete one cycle of stratified ranked set sample. The cycle may be repeated m times until $n = mr$ elements have been measured (see Samawi, 1996).

The following stricture for the stratified Ranked set sample is used when the ranking on the variable *Y* in case of bivariate elements: For the *k-th* cycle, the SRSS is denoted by

Stratum 1: $\left(Y_{1(1)k}, X_{1[1]k}\right), \left(Y_{1(2)k}, X_{1[2]k}\right), ..., \left(Y_{1(r_1)k}, X_{1[r_1]k}\right)$

Stratum 2: $\left(Y_{2(1)k}, X_{2[1]k}\right), \left(Y_{2(2)k}, X_{2[2]k}\right), ..., \left(Y_{2(r_2)k}, X_{2[r_2]k}\right)$

$\vdots$

Stratum L: $\left(Y_{L(1)k}, X_{L[1]k}\right), \left(Y_{L(2)k}, X_{L[2]k}\right), ..., \left(Y_{L(r_L)k}, X_{L[r_L]k}\right), k = 1, 2, ..., m$

Similarly for the stratified ranked set sample when the ranking on the variable *X*:

Stratum 1: $\left(Y_{1[1]k}, X_{1(1)k}\right), \left(Y_{1[2]k}, X_{1(2)k}\right), \ldots, \left(Y_{1[r_1]k}, X_{1(r_1)k}\right)$

Stratum 2: $\left(Y_{2[1]k}, X_{2(1)k}\right), \left(Y_{2[2]k}, X_{2(2)k}\right), \ldots, \left(Y_{2[r_2]k}, X_{2(r_2)k}\right)$

$\vdots$

Stratum L: $\left(Y_{L[1]k}, X_{L(1)k}\right), \left(Y_{L[2]k}, X_{L(2)k}\right), \ldots, \left(Y_{L[r_L]k}, X_{L(r_L)k}\right)$

where $k = 1,2,\ldots,m$.

Extreme Ranked Set Sample

The extreme ranked set sample ERSSs investigated by Samawi et al. (1996). The procedure involves randomly drawing *r* sets of *r* units each, from the infinite population for which the mean is to be estimated. It is assumed that the lowest or the largest units of this set can be detected visually or with little cost. For sure, this is a simple and practical process. From the fist set of *r* units the lowest ranked unit is measured. From the second set of *r* units the largest ranked unit is measured. From the third set of *r* units the lowest ranked unit is measured, and so on. In this way we obtain the first (*r*-1) measured units using the first (*r* -1) sets. The choice of the *r- th* unit from the *r-th* (i.e the last) set depends on whether *r* is even or odd.

a) If *r* is even the largest ranked unit is measured. ERSSa will denote such a sample.

If *r* is odd then two options exist:

b) For the measure of the *r-th* unit we take the average of the measures of the lowest and the largest units in the *r-th* set. ERSSb will denote such a sample.

c) For the measure of the *r*-th unit we take the measure of the median. ERSSc will denote such a sample. Note that the choice (c) will be more difficult in application than the choice (a) and (b).

Stratified Extreme Ranked Set Sample

Suppose that the population divided into L mutually exclusive and exhaustive strata, with subpopulation size $N_1, N_2, \ldots, N_L$. Through this article it large subpopulation and symmetric underlying distribution will be assumed. The following notations and results will be introduced for this paper. For all ($i = 1, 2, \ldots, r_h$) and $h = 1, 2, \ldots, L$.

Let $\mu_{Xh} = E\left(X_{hij}\right), \sigma^2_{Xh} = Var\left(X_{hij}\right),$

$\mu_{Xh(i)} = E\left(X_{hi(i)}\right), \sigma^2_{Xh(i)} = Var\left(X_{hi(i)}\right)$ and

$W_h = \dfrac{N_h}{N} = \dfrac{n_h}{n} = \dfrac{r_h}{r}$ (proportional allocation).

Let $X^*_{h11}, X^*_{h12}, \ldots, X^*_{h1r_h};$

$X^*_{h21}, X^*_{h22}, \ldots, X^*_{h2r_h}; \ldots;$

$X^*_{hr_h1}, X^*_{hr_h1}, \ldots, X^*_{hr_hr_h}$ be r_h independent samples of size r_h, each taken from the $h-th$ stratum ($h = 1, 2, \ldots, L$). Assume that each element X^*_{hij} in the sample has the same distribution function $F_h(x)$ with mean μ_{Xh} and variance σ^2_{Xh}. For simplicity of notation, we will assume that X_{hij} denotes the quantitative measure of the unit X^*_{hij}.

Then, according to our description $X_{h11}, X_{h21}, \ldots, X_{hr_h1}$ is the SRS from the $h-th$ stratum. Let $X^*_{hi(1)}, X^*_{hi(2)}, \ldots, X^*_{hi(r_h)}$ be the ordered statistics of the $i-th$ sample $X^*_{hi1}, X^*_{hi2}, \ldots, X^*_{hir_h}$, ($i = 1, 2, \ldots, r_h$), taken from the $h-th$ stratum. If r_h is even then $X_{h1(1)}, X_{h2(r_h)}, X_{h3(1)}, \ldots, X_{h\{r_h-1\}(1)}, X_{hr_h(r_h)}$ denotes the ERSSh a for the $h-th$ stratum. If

r_h is odd then

$X_{h1(1)},\ X_{h2(r_h)},\ X_{h3(1)},\ \dots,\ X_{h\{r_h-1\}(r_h)},\ X_{hr_h\left(\frac{r_h+1}{2}\right)}$

denotes the ERSSh c for the $h-th$ stratum. Note that this will be repeated for each ($h=1,2,\dots,L$). The resulting L independent ERSSs from each stratum will be denotes the stratified extreme ranked set sample SERSS. This process can be repeated m independent times.

Estimate of Population Mean Using

To estimate the mean μ using SERSS of size n, assume that there is (a) strata with even set size and (L-a) strata with odd set size. For simplicity of notation, let m=1 then $n_h = r_h$ and $n=r$, then the estimate of the mean μ_X using SERSS is given by

$$\overline{X}_{SERSS} = \sum_{h=1}^{a} W_h \overline{X}_{h(a)} + \sum_{h=a+1}^{L} W_h \overline{X}_{h(c)}, \text{ where}$$

$$\overline{X}_{h(a)} = \frac{1}{2}\left(\overline{X}_{h(1)} + \overline{X}_{h(r_h)}\right),$$

$$\overline{X}_{h(c)} = \frac{X_{h1(1)} + X_{h2(r_h)} + X_{h3(1)} + \dots + X_{h\{r_h-1\}}(r_h) + X_{hr_h\left(\frac{r_h+1}{2}\right)}}{r_h},$$

$$\overline{X}_{h(1)} = 2\sum_{i=1}^{\frac{r_h}{2}} \frac{X_{h\{2i-1\}(1)}}{r_h} \quad \text{and}$$

$$\overline{X}_{h(r_h)} = 2\sum_{i=1}^{\frac{r_h}{2}} \frac{X_{h2i(r_h)}}{r_h}.$$

It can be shown that (Samawi et al., 1996).

$$E\left(\overline{X}_{h(a)}\right) = \frac{1}{2}\left(\mu_{Xh(1)} + \mu_{Xh(r_h)}\right) \quad \text{and}$$

$$E\left(\overline{X}_{h(c)}\right) = \left(\frac{r_h-1}{2r_h}\right)\left(\mu_{Xh(1)} + \mu_{Xh(r_h)}\right) + \frac{1}{r_h}\mu_{Xh\left(\frac{r_h+1}{2}\right)}.$$

Therefore, the mean and variance of $\overline{X}_{SERSS}$ are

$$E(\overline{X}_{SERSS}) = \frac{1}{2}\sum_{h=1}^{L} W_h\left(\mu_{Xh(1)} + \mu_{Xh(r_h)}\right), + \frac{1}{2}\sum_{h=a+1}^{L}\frac{W_h}{r_h}\left[2\mu_{Xh\left(\frac{r_h+1}{2}\right)} - \left(\mu_{Xh(1)} + \mu_{Xh(r_h)}\right)\right]$$

and

$$Var\left(\overline{X}_{SERSS}\right) = \frac{1}{2}\sum_{h=1}^{L}\frac{W_h^2}{r_h}\left(\sigma^2_{Xh(1)} + \sigma^2_{Xh(r_h)}\right) + \frac{1}{2}\sum_{h=a+1}^{L}\frac{W_h^2}{r_h}\left[2\sigma^2_{Xh\left(\frac{r_h+1}{2}\right)} - \left(\sigma^2_{Xh(1)} + \sigma^2_{Xh(r_h)}\right)\right].$$

Note that the elements in $\overline{X}_{h(a)}$ are independent and so are the elements in $\overline{X}_{h(c)}$. Furthermore, the elements in $\overline{X}_{h(a)}$ are independent of the elements in $\overline{X}_{h(c)}$ and so are independent of the element in $X_{h\left(\frac{r_h+1}{2}\right)}$. If the underlying distribution for each stratum is symmetric then it can be shown that $E\left(\tilde{X}_{SERSS}\right) = \mu_X$ (i.e., an unbiased estimator) and

$$Var\left(\overline{X}_{SERSS}\right) = \sum_{h=1}^{L}\frac{W_h^2}{r_h}\sigma^2_{Xh(1)} + \sum_{h=a+1}^{L}\frac{W_h^2}{r_h^2}\left[\sigma^2_{Xh\left(\frac{r_h+1}{2}\right)} - \sigma^2_{Xh(1)}\right] \tag{2.1}$$

Note that the estimate of the mean μ using SSRS of size r is given by $\overline{X}_{SSRS} = \sum_{h=1}^{L} W_h \overline{X}_h$. Also, the mean and variance of $\overline{X}_{SSRS}$ are known to be $E\left(\overline{X}_{SSRS}\right) = \mu_X$ (i.e., an unbiased estimator) and

$$Var\left(\overline{X}_{SSRS}\right)=\sum_{h=1}^{L}W_h^2\,\frac{\sigma_{Xh}^2}{r_h} \qquad (2.2)$$

(see Cochran,1977).

Theorem: Assume that the underlying distribution for each stratum follows Normal or Logistic distribution. Then $Var\left(\overline{X}_{SERSS}\right)\le Var\left(\overline{X}_{SSRS}\right)$.

Proof: Assume large subpopulation sizes ($N_1, N_2, \cdots, N_L$). In case of Normal, or Logistic distribution functions the following are true, $\sigma^2_{Xh(1)} \ge \sigma^2_{Xh(2)} \ge, \dots, \ge \sigma^2_{Xh\left(\frac{r_h}{2}\right)}$ if r_h is even and $\sigma^2_{Xh(1)} \ge \sigma^2_{Xh(2)} \ge, \dots, \ge \sigma^2_{Xh\left(\frac{r_h+1}{2}\right)}$ if r_h is odd.

Also note the $\sigma^2_{Xh(i)} \le \sigma^2_{Xh}$, $i=1, 2, \dots, r_h$ (Arnold, 1992.) By comparing (2.1) and (2.2), and since $\sigma^2_{Xh(1)} \le \sigma^2_{Xh}, \sigma^2_{Xh\left(\frac{r_h+1}{2}\right)} \le \sigma^2_{Xh(1)}$ and

$\sigma^2_{Xh\left(\frac{r_h+1}{2}\right)} - \sigma^2_{Xh(1)} \le 0$, therefore

$\sigma^2_{Xh\left(\frac{r_h+1}{2}\right)} - \sigma^2_{Xh(1)} \le 0$, therefore

$$\sum_{h=1}^{L}\frac{W_h^2}{r_h}\sigma^2_{Xh(1)} + \sum_{h=a+1}^{L}\frac{W_h^2}{r_h^2}\left[\sigma^2_{Xh\left(\frac{r_h+1}{2}\right)} - \sigma^2_{Xh(1)}\right]$$

$$\le \sum_{h=1}^{L}W_h^2\,\frac{\sigma^2_{Xh}}{r_h},$$

and hence
$Var\left(\overline{X}_{SERSS}\right)\le Var\left(\overline{X}_{SSRS}\right)$.

Simulation Study

The normal and logistic distribution is used in the simulation. Sample size r = 10, 20 and 30 and number of strata L = 3 are considered. For each of the possible combination of distribution, sample size and different choice of parameters 2000 data sets were generated. The relative efficiencies of the estimate of the population mean using SERSS with respect to SSRS, SRS, and RSS are obtained.

The values obtained by simulation are given in Table 1. Our Simulation indicates that estimating the population means using SERSS is more efficient than using SSRS or SRS. In some case, when the underling distribution is normal with ($\mu_1 = 1.0, \mu_2 = 3.0, \mu_3 = 5.0, r = 10$), the simulation indicates that estimating the population mean using SERSS is even more efficient than using RSS, of the same size.

Separate Ratio Estimation using SERSS

In this Section, obtain the separate ratio estimator was obtained using stratified extreme ranked set sample. Also, the asymptotic mean and variance of the estimator were derived. Two cases are considered, the first case if the ranking on variable Y is perfect, while the ranking of the variable X will be with errors in ranking. The second case, when the ranking on variable X is perfect, while the ranking of the variable Y will be with errors in ranking. Also, some comparisons of the two cases are investigated.

Table 1. The relative efficiency of the simulation results.

Distribution function	n	$RE(\overline{X}_{SSRS}, \overline{X}_{SSRSS})$	$RE(\overline{X}_{RSS}, \overline{X}_{SERSS})$	$RE(\overline{X}_{SRS}, \overline{X}_{SERSS})$
Normal				
$W_1 = 0.3, W_2 = 0.3, W_3 = 0.4$	10	1.20	1.43	7.42
$\mu_1 = 1.0, \mu_2 = 3.0, \mu_3 = 5.0$	20	2.34	0.94	8.47
	30	2.80	0.75	9.91
$\sigma_1 = 1.0, \sigma_2 = 1.0, \sigma_3 = 1.0$				
Normal				
$W_1 = 0.3, W_2 = 0.3, W_3 = 0.4$	10	2.07	0.74	3.46
$\mu_1 = 1.0, \mu_2 = 2.0, \mu_3 = 3.0$	20	2.39	0.53	4.04
$\sigma_1 = 1.0, \sigma_2 = 1.0, \sigma_3 = 1.0$	30	2.77	0.42	4.70
Normal				
$W_1 = 0.3, W_2 = 0.3, W_3 = 0.4$	10	1.92	1.21	6.17
$\mu_1 = 1.0, \mu_2 = 3.0, \mu_3 = 5.0$	20	2.29	0.80	7.41
	30	2.73	0.67	9.05
$\sigma_1 = 1.0, \sigma_2 = 1.1, \sigma_3 = 1.2$				
Logtic				
$W_1 = 0.3, W_2 = 0.3, W_3 = 0.4$	10	1.72	0.71	3.14
$\mu_1 = 1.0, \mu_2 = 3.0, \mu_3 = 5.0$	20	1.74	0.43	3.20
	30	1.87	0.35	3.56
$\sigma_1 = 1.0, \sigma_2 = 1.0, \sigma_3 = 1.0$				
Logtic				
$W_1 = 0.3, W_2 = 0.3, W_3 = 0.4$	10	1.93	0.51	2.19
$\mu_1 = 1.0, \mu_2 = 2.0, \mu_3 = 3.0$	20	1.78	0.30	1.98
	30	1.89	0.29	2.26
$\sigma_1 = 1.0, \sigma_2 = 1.0, \sigma_3 = 1.0$				
Logtic				
$W_1 = 0.3, W_2 = 0.3, W_3 = 0.4$	10	1.79	0.68	3.26
$\mu_1 = 1.0, \mu_2 = 3.0, \mu_3 = 5.0$	20	1.78	0.40	3.09
$\sigma_1 = 1.0, \sigma_2 = 1.1, \sigma_3 = 1.2$	30	1.83	0.34	3.42

Ratio Estimation when ranking on Variable *Y*.

Assuming that we can only rank on the variable *Y* so that the ranking of *Y* will be perfect while the ranking of *X* will be with error in ranking. If r_h is even then

$$\left(X_{h1[1]k}, Y_{h1(1)k}\right), \left(X_{h2[r_h]k}, Y_{h2(r_h)k}\right), \left(X_{h3[1]k}, Y_{h3(1)k}\right), \ldots, \left(X_{h\{r_h-1\}[1]k}, Y_{h\{r_h-1\}(1)k}\right), \left(X_{hr_h[r_h]k}, Y_{hr_h(r_h)k}\right)$$

denotes the SERSS1$_a$ for the *h-th* stratum. If r_h is odd then

$$\left(X_{h1[1]k}, Y_{h1(1)k}\right), \left(X_{h2[r_h]k}, Y_{h2(r_h)k}\right), \left(X_{h3[1]k}, Y_{h3(1)k}\right), \ldots, \left(X_{h\{r_h-1\}(r_h)}, Y_{h\{r_h-1\}(r_h)}\right), \left(X_{hr_h\left[\frac{r_h+1}{2}\right]k}, Y_{hr_h\left(\frac{r_h+1}{2}\right)k}\right)$$

denotes the SERSS1$_c$ for the *h-th* stratum, *k*=1, 2, ... , *m*.

The separate ratio estimate requires knowledge of the stratum totals η_h in order to be used for estimating the population mean or total. Then using the same notation of Section (2.1) of the SERSS when ranking on variable Y, then the ratio can be estimated within each stratum as follows:

$$\hat{R}_{SERSSh1a} = \frac{\overline{Y}_{h(a)}}{\overline{X}_{h[a]}} \quad \text{if } (r_h) \text{ is even}$$

and $$\hat{R}_{SERSSh1c} = \frac{\overline{Y}_{h(c)}}{\overline{X}_{h[c]}} \quad \text{if } (r_h) \text{ is odd}$$

where

$$\overline{X}_{h[a]} = \frac{1}{2}\left(\overline{X}_{h[1]} + \overline{X}_{h[r_h]}\right)$$

$$\overline{X}_{h[1]} = 2\sum_{k=1}^{m}\sum_{i=1}^{\frac{r_h}{2}} \frac{X_{h\{2i-1\}[1]k}}{mr_h},$$

$$\overline{X}_{h[r_h]} = 2\sum_{k=1}^{m}\sum_{i=1}^{\frac{r_h}{2}} \frac{X_{h2i[r_h]k}}{mr_h},$$

$$\overline{X}_{h[c]} = \frac{\sum_{k=1}^{m}\left\{ X_{h1[1]k} + X_{h2[r_r]k} + X_{h3[1]k} + \ldots + X_{h\{r_h-1\}[r_h]} + X_{hr_h\left[\frac{r_h+1}{2}\right]k} \right\}}{mr_h},$$

$$\overline{Y}_{h(a)} = \frac{1}{2}\left(\overline{Y}_{h(1)} + \overline{Y}_{h(r_h)}\right),$$

$$\overline{Y}_{h(1)} = 2\sum_{k=1}^{m}\sum_{i=1}^{\frac{r_h}{2}} \frac{Y_{h\{2i-1\}(1)k}}{mr_h},$$

$$\overline{Y}_{h(r_h)} = 2\sum_{k=1}^{m}\sum_{i=1}^{\frac{r_h}{2}} \frac{Y_{h2i(r_h)k}}{mr_h} \quad \text{and } n_h = mr_h,$$

and

$$\overline{Y}_{h(c)} = \frac{\sum_{k=1}^{m}\left\{ Y_{h1(1)k} + Y_{h2(r_h)k} + Y_{h3(1)k} + \ldots + Y_{h\{r_h-1\}[r_h]} + Y_{hr_h\left[\frac{r_h+1}{2}\right]k} \right\}}{mr_h}$$

Note that the sample sizes are different from one stratum to another. Therefore, assume without loss of generality that the first (*a*) strata have even set size (r_h), $h = 1, 2, \ldots, a$, and the last (*L-a*) strata have odd set size (r_h), $h = a+1, a+2, \ldots, L$. This implies that, the separate ratio estimator using stratified extreme ranked set sample when the ranking on variable *Y*, will be as follows:

$$\hat{R}_{SERSS1} = \sum_{h=1}^{a}\frac{\eta_h}{\eta}\frac{\overline{Y}_{h(a)}}{\overline{X}_{h[a]}} + \sum_{h=a+1}^{L}\frac{\eta_h}{\eta}\frac{\overline{Y}_{h(c)}}{\overline{X}_{h[c]}}, \tag{3.1}$$

$$\hat{R}_{SERSS1} = \sum_{h=1}^{a} W_h \frac{\mu_{Xh}}{\mu_X}\frac{\overline{Y}_{h(a)}}{\overline{X}_{h[a]}} + \sum_{h=a+1}^{L} W_h \frac{\mu_{Xh}}{\mu_X}\frac{\overline{Y}_{h(c)}}{\overline{X}_{h[c]}}\,, \quad (3.2)$$

where

$W_h = \frac{N_h}{N}$, $\eta_h = N_h \mu_{Xh}$ and $\eta = N\mu_X$ (known).

It can be shown using the Taylor series expansion method that

$$E(\hat{R}_{SERSS1}) = \frac{\mu_Y}{\mu_X} + O\left(\min_h \left(mr_h\right)^{-1}\right).$$

Also, the approximate variance of $\hat{R}_{SERSS1}$ can be obtain as follows: Since we have independent strata and the assumption of symmetric marginal distribution, then

$$Var(\hat{R}_{SERSS1}) = Var\left(\sum_{h=1}^{a} W_h \frac{\mu_{Xh}}{\mu_X}\frac{\overline{Y}_{h(a)}}{\overline{X}_{h[a]}} + \sum_{h=a+1}^{L} W_h \frac{\mu_{Xh}}{\mu_X}\frac{\overline{Y}_{h(c)}}{\overline{X}_{h[c]}}\right)$$

$$Var(\hat{R}_{SERSS1}) = \sum_{h=1}^{a} W_h^2 \frac{\mu_{Xh}^2}{\mu_X^2} V(\hat{R}_{ERSSh1a}) + \sum_{h=a+1}^{L} W_h^2 \frac{\mu_{Xh}^2}{\mu_X^2} V(\hat{R}_{ERSSh1c}) \quad (3.3)$$

Using similar argument as in Samawi and Muttlak (1996), we have

$$Var(\hat{R}_{ERSSh1a}) = \frac{R_h^2}{r_h m}\left(\frac{\sigma^2_{Xh[1]}}{\mu^2_{Xh}} + \frac{\sigma^2_{Yh(1)}}{\mu^2_{Yh}} - 2\frac{\sigma_{Xh[1]Yh(1)}}{\mu_{Xh}\mu_{Yh}}\right) \quad (3.4)$$

$$-2\frac{(r_h-1)\sigma_{Xh[1]Yh(1)} + \sigma_{Xh\left[\frac{r_h+1}{2}\right]Yh\left(\frac{r_h+1}{2}\right)}}{r_h \mu_{Xh}\mu_{Yh}}) \quad (3.5)$$

Therefore, the approximate variance of separate ratio estimator using SERSS (ranking on variable Y) is

$$Var(\hat{R}_{ERSSh1c}) = \frac{R_h^2}{n_h}\left(\frac{(r_h-1)\sigma^2_{Xh[1]} + \sigma^2_{X[1]h\left[\frac{r+1}{2}\right]}}{r_h\mu^2_{Xh}} + \frac{(r_h-1)\sigma^2_{Yh(1)} + \sigma^2_{Yh\left(\frac{r_h+1}{2}\right)}}{r_h\mu^2_{Yh}}\right.$$

$$Var(\hat{R}_{SERSS1}) = \sum_{h=1}^{L} W_h^2 \frac{\mu^2_{Xh}}{\mu^2_X}\frac{R_h^2}{n_h}\left(\frac{\sigma^2_{Xh[1]}}{\mu^2_{Xh}} + \frac{\sigma^2_{Yh(1)}}{\mu^2_{Yh}} - 2\frac{\sigma_{Xh[1]Yh(1)}}{\mu_{Xh}\mu_{Yh}}\right)$$

$$+\sum_{h=a+1}^{L} W_h^2 \frac{\mu^2_{Xh}}{\mu^2_X}\frac{R_h^2}{n_h}\left(\frac{\sigma^2_{Xh\left[\frac{r_h+1}{2}\right]} - \sigma^2_{Xh[1]}}{r_h\mu^2_{Xh}} + \frac{\sigma^2_{Yh\left(\frac{r_h+1}{2}\right)} - \sigma^2_{Yh(1)}}{r_h\mu^2_{Yh}} - 2\frac{\sigma_{Xh\left[\frac{r_h+1}{2}\right]Yh\left(\frac{r_h+1}{2}\right)} - \sigma_{Xh[1]Yh(1)}}{r_h\mu_{Xh}\mu_{Yh}}\right), \quad (3.6)$$

where $\frac{\sigma^2_{Xh[1]}}{\mu^2_{Xh}} = \frac{E\left(X_{h[1]} - \mu_{Xh[1]}\right)^2}{E\left(X_h\right)^2}$,

$$\frac{\sigma^2_{Yh(1)}}{\mu^2_{Yh}} = \frac{E\left(Y_{h(1)} - \mu_{Yh(1)}\right)^2}{E\left(Y_h\right)^2}$$

$$\frac{\sigma^2_{Xh\left[\frac{r_h+1}{2}\right]}}{\mu^2_{Xh}} = \frac{E\left(X_{h\left[\frac{r_r+1}{2}\right]} - \mu_{Xh\left[\frac{r_r+1}{2}\right]}\right)^2}{E\left(X_h\right)^2},$$

$$\frac{\sigma^2_{Yh\left(\frac{r_h+1}{2}\right)}}{\mu^2_{Yh}} = \frac{E\left(Y_{h\left(\frac{r_h+1}{2}\right)} - \mu_{Yh\left(\frac{r_h+1}{2}\right)}\right)^2}{E\left(Y_h\right)^2}$$

$$\sigma_{Xh[1]Yh(1)} = E\left(Y_{h(1)} - \mu_{Yh(1)}\right)\left(X_{h[1]} - \mu_{Xh[1]}\right)$$

and $R_h = \dfrac{\mu_{Yh}}{\mu_{Xh}}$.

Ratio Estimation when Ranking on Variable X.

Similarly by changing the notation of perfect ranking ( ), by imperfect ranking [ ], for *X* and *Y*. Also, by using the same notation of the SERSS when ranking on variable *X*, then the separate ratio estimator using SERSS when the ranking on variable *X*, will be as follows:

$$\hat{R}_{SERSS2} = \sum_{h=1}^{a} W_h \frac{\mu_{Xh}}{\mu_X} \frac{\overline{Y}_{h[a]}}{\overline{X}_{h(a)}} + \sum_{h=a+1}^{L} W_h \frac{\mu_{Xh}}{\mu_X} \frac{\overline{Y}_{h[c]}}{\overline{X}_{h(c)}} \quad (3.7)$$

In the same way as in section (3.1) we get the following results:

$$Var(\hat{R}_{SERSS2}) = \sum_{h=1}^{L} W_h^2 \frac{\mu^2_{Xh}}{\mu^2_X} \frac{R_h^2}{n_h} \left(\frac{\sigma^2_{Xh(1)}}{\mu^2_{Xh}} + \frac{\sigma^2_{Yh[1]}}{\mu^2_{Yh}} - 2\frac{\sigma_{Xh(1)Yh[1]}}{\mu_{Xh}\mu_{Yh}}\right)$$
$$+ \sum_{h=a+1}^{L} W_h^2 \frac{\mu^2_{Xh}}{\mu^2_X} \frac{R_h^2}{n_h} \left(\frac{\sigma^2_{Xh\left(\frac{r_h+1}{2}\right)} - \sigma^2_{Xh(1)}}{r_h\mu^2_{Xh}} + \frac{\sigma^2_{Yh\left[\frac{r_h+1}{2}\right]} - \sigma^2_{Yh[1]}}{r_h\mu^2_{Yh}} - 2\frac{\sigma_{Xh\left(\frac{r_h+1}{2}\right)Yh\left[\frac{r_h+1}{2}\right]} - \sigma_{Xh(1)Yh[1]}}{r_h\mu_{Xh}\mu_{Yh}}\right). \quad (3.8)$$

Ranking on which variable?

Again, since one can not rank on both variables at the same time and some time it is easier to rank on one variable than the other, then we need to decide on which variable we should rank. We need to compare the variance of $\hat{R}_{SERSS1}$ in (3.6) and variance $\hat{R}_{SERSS2}$

Theorem 3.2 : Assume that there are L linear relations between Y_h and X_h, i.e., $|\rho_h| > 0$ and it is easy to rank on variable X. Also assume that the approximation to the variance of the ratio estimators $\hat{R}_{SERSS1}$ and $\hat{R}_{SERSS2}$ given in equations (3.6) and (3.8) respectively are valid and the bias of the estimators can be ignored. If underlying distribution are Normal or Logistic distribution, then $Var(\hat{R}_{SERSS2}) \le Var(\hat{R}_{SERSS1})$.

Proof : To prove the above we consider simple linear regression model between Y_h and X_h, and Y & X each has either Normal or Logistic marginal distribution function.

$$Y_{hi} = \alpha_h + \beta_h X_{hi} + \varepsilon_{hi}, \quad (3.9)$$

$$\mu_{Yh} = \alpha_h + \beta_h \mu_{Xh} \quad (3.10)$$

where α_h and β_h are parameters and ε_{hi} is a random error with $E(\varepsilon_{hi}) = 0$, $Var(\varepsilon_{hi}) = \sigma^2_h$ and $Cov(\varepsilon_{hi}, \varepsilon_{hj}) = 0$ for $i \ne j$, $i = 1, 2, \ldots, r_h$, also ε_{hi} and X_{hi} are independent. Let

$$\sigma^2 Xh[i] = \begin{cases} \sigma^2_{Xh(i)} & \text{if the ranking of the i-th order statistic in the i-th sample is correct} \\ \sigma^2_{Xh} & \text{if the ranking of the i-th order statistic in the i-th sample is not correct i.e., radom order} \end{cases}$$

Note that, according to our definition and by the assumption of the underlying distributions

$\sigma^2_{Xh(i)} \le \sigma^2_{Xh[i]}$ (Arnold, 1992).

Case 1. If we are ranking on the Y_h variable we get the following model from equation (3.9)

$$Y_{h(i)} = \alpha_h + \beta_h X_{h[i]} + \varepsilon_{h[i]}, \qquad (3.11)$$

where $\varepsilon_{h[i]}$ is a random error with $E(\varepsilon_{h[i]}) = 0$, $Var(\varepsilon_{h[i]}) = \sigma^2_h$ and $Cov(\varepsilon_{h[i]}, \varepsilon_{h[j]}) = 0$ for $i \ne j,\ i = 1, 2, \dots, r_h$ also $\varepsilon_{h[i]}$ and $X_{h[i]}$ are independent. The expected value of $Y_{h(i)}$ can be written as

$$\mu_{Yh(i)} = \alpha_h + \beta_h \mu_{Xh[i]}. \qquad (3.12)$$

The variance of $Y_{h(i)}$ is

$$\sigma^2_{Yh(i)} = \beta^2_h \sigma^2_{Xh[i]} + \sigma^2_h. \qquad (3.13)$$

Now, by subtracting $\mu_{Yh(i)}$ from equation (3.11) and multiply both sides by $(X_{h[i]} - \mu_{Xh[i]})$, and then take the expected value for the both sides we get,

$$\sigma_{Yh(i)Xh[i]} = \beta_h \sigma^2_{Xh[i]} \qquad (3.14)$$

Case 2. If we are ranking on the variable *X* we get the following model from equation (3.9)

$$Y_{h[i]} = \alpha_h + \beta_h X_{h(i)} + \varepsilon_{h[i]}. \qquad (3.15)$$

The expected value of $Y_{h[i]}$ is

$$\mu_{Yh[i]} = \alpha_h + \beta_h \mu_{Xh(i)}. \qquad (3.16)$$

Similarly we can show that

$$\sigma^2_{Yh[i]} = \beta^2_h \sigma^2_{Xh(i)} + \sigma^2_h. \qquad (3.17)$$

and $\sigma_{Yh[i]Xh(i)} = \beta_h \sigma^2_{Xh(i)}$ (3.18)

Now, equations (3.6) and (3.8) can be writen as :

$$Var(\hat{R}_{SERSS1}) = \sum_{h=1}^{L} W_h^2 \frac{\mu^2_{Xh}}{\mu^2_X} \frac{R^2_h}{n_h} \left(\sigma^2_{Xh[1]} \frac{(\mu_{Yh} - \beta_h \mu_{Yh})^2}{\mu^2_{Yh}\mu^2_{Xh}} + \frac{\sigma^2_h}{\mu^2_{Yh}}\right)$$
$$+ \sum_{h=a+1}^{L} W_h^2 \frac{\mu^2_{Xh}}{\mu^2_x} \frac{R^2_h}{m r^2_h} \left\{ \left[\sigma^2_{Xh\left[\frac{r_h+1}{2}\right]} - \sigma^2_{Xh[1]} \right] \frac{\left(\mu_{Yh} - \beta_h \mu_{X\ h}\right)^2}{\mu^2_{Yh}\ \mu^2_{xh}} \right\}$$

and

$$Var(\hat{R}_{SERSS2}) = \sum_{h=1}^{L} W_h^2 \frac{\mu^2_{Xh}}{\mu^2_X} \frac{R^2_h}{n_h} \left(\sigma^2_{Xh(1)} \frac{(\mu_{Yh} - \beta_h \mu_{Yh})^2}{\mu^2_{Yh}\mu^2_{Xh}} + \frac{\sigma^2_h}{\mu^2_{Yh}}\right)$$
$$+ \sum_{h=a+1}^{L} W_h^2 \frac{\mu^2_{Xh}}{\mu^2_x} \frac{R^2_h}{m r^2_h} \left\{ \left[\sigma^2_{Xh\left(\frac{r_h+1}{2}\right)} - \sigma^2_{Xh(1)} \right] \frac{\left(\mu_{Yh} - \beta_h \mu_{X\ h}\right)^2}{\mu^2_{Yh}\ \mu^2_{xh}} \right\}$$

respectively, therefore

$Var(\hat{R}_{SERSS2}) \le Var(\hat{R}_{SERSS1})$.

Finally in this case it is recommended to rank on variable that will be used in the denominator of the ratio estimator if we wish to estimate the mean or the total of the population using the ratio estimator method when the data is selected according SERSS method.

Combined Ratio Estimation using SERSS

In this Section, combined were ratio estimator using stratified extreme ranked set sample. Two cases were considered, the first is to make the ranking on variable Y perfect, while the ranking on the variable X will be with errors in ranking. The second case, when the ranking on variable X is perfect, while the ranking on the variable Y will be with errors in ranking. Also, the properties of these estimators will be discussed.

Using SERSS as describe in Section 2, when ranking on variable *Y*. The combined Ratio estimate is defined by:

$$\hat{R}_{SERSS1(s)} = \frac{\bar{Y}_{(SERSS)}}{\bar{X}_{[SERSS]}}, \qquad (4.1)$$

where

$$\bar{Y}_{(SERSS)} = \sum_{h=1}^{a} W_h \bar{Y}_{h(a)} + \sum_{h=a+1}^{L} W_h \bar{Y}_{h(c)},$$

$$\bar{X}_{[SERSS]} = \sum_{h=1}^{a} W_h \bar{X}_{h[a]} + \sum_{h=a+1}^{L} W_h \bar{X}_{h[c]}.$$

Therefore,

$$\hat{R}_{SERSS1} = \frac{\sum_{h=1}^{a} W_h \bar{Y}_{h(a)} + \sum_{h=a+1}^{L} W_h \bar{Y}_{h(c)}}{\sum_{h=1}^{a} W_h \bar{X}_{h[a]} + \sum_{h=a+1}^{L} W_h \bar{X}_{h[c]}}. \qquad (4.2)$$

For fixed r_h, assume that we have finite second moments for *X* and *Y*. Since the ratio is a function of the means of X and Y, i.e., $R = \frac{\mu_Y}{\mu_X}$, and hence R has at least two bounded derivations of all types in some neighborhood of (μ_Y, μ_X) provided that $\mu_X \neq 0$. Then, assuming large m, we can use the Multivariate Taylor Series Expansion, to approximate the variance and get the order of the bias of the ratio estimator. Therefore,

$$Var(\hat{R}_{SERSS1}) = R^2 \sum_{h=1}^{a} \frac{W_h^2}{mr_h} \left(\frac{\sigma^2_{Xh[1]}}{\mu_X^2} + \frac{\sigma^2_{Yh(1)}}{\mu_Y^2} - 2\frac{\sigma_{Xh[1]Yh(1)}}{\mu_X \mu_Y} \right)$$

$$+R^2 \sum_{h=a+1}^{L} \frac{W_h^2}{mr_h} \left(\frac{(r_h - 1)\sigma^2_{Xh[1]} + \sigma^2_{Xh\left[\frac{r_h+1}{2}\right]}}{r_h \mu_X^2} + \frac{(r_h - 1)\sigma^2_{Yh(1)} + \sigma^2_{Yh\left(\frac{r_h+1}{2}\right)}}{r_h \mu_Y^2} - 2\frac{(r_h - 1)\sigma_{Xh[1]Yh(1)} + \sigma_{Xh\left[\frac{r_h+1}{2}\right]Yh\left(\frac{r_h+1}{2}\right)}}{r_h \mu_X \mu_Y} \right). \qquad (4.3)$$

Ratio Estimation when Ranking on variable X.

Similarly, the estimate is given by:

$$\hat{R}_{SERSS} = \frac{\bar{Y}[SERSS]}{\bar{X}(SERSS)}, \qquad (4.4)$$

where

$$\bar{Y}_{(SERSS)} = \sum_{h=1}^{a} W_h \bar{Y}_{h(a)} + \sum_{h=a+1}^{L} W_h \bar{Y}_{h(c)}$$

$$\bar{X}_{[SERSS]} = \sum_{h=1}^{a} W_h \bar{X}_{h[a]} + \sum_{h=a+1}^{L} W_h \bar{X}_{h[c]}$$

Therefore, in combined case, we get:

$$\hat{R}_{SERSS1} = \frac{\sum_{h=1}^{a} W_h \bar{Y}_{h[a]} + \sum_{h=a+1}^{L} W_h \bar{Y}_{h[c]}}{\sum_{h=1}^{a} W_h \bar{X}_{h(a)} + \sum_{h=a+1}^{L} W_h \bar{X}_{h(c)}} \qquad (4.5)$$

Using the same argument as in section (4.1), $E\left(\hat{R}_{SERSSS2}\right) \approx R + O\left(\min_h (mr_h)^{-1}\right)$, and

$$Var(\hat{R}_{SERSS\,2})$$
$$= R^2 \sum_{h=1}^{a} \frac{W_h^2}{mr_h} \left(\frac{\sigma^2_{Xh(1)}}{\mu_X^2} + \frac{\sigma^2_{Yh[1]}}{\mu_Y^2} - 2\frac{\sigma_{Xh(1)Yh[1]}}{\mu_X \mu_Y} \right)$$

$$+R^2 \sum_{h=a+1}^{L} \frac{W_h^2}{mr_h} \left(\frac{(r_h-1)\sigma^2_{Xh(1)} + \sigma^2_{Xh\left(\frac{r_h+1}{2}\right)}}{r_h \mu_h^2} \right.$$
$$+ \frac{(r_{h-1})\sigma^2_{Yh[1]} + \sigma^2_{Yh\left[\frac{r_h+1}{2}\right]}}{r_h \mu_Y^2}$$

$$\left. - 2\frac{(r_h-1)\sigma_{Xh[1]Yh(1)} + \sigma_{Xh\left[\frac{r_h+1}{2}\right]Yh\left(\frac{r_h+1}{2}\right)}}{r_h \mu_X \mu_Y} \right). \quad (4.6)$$

Ranking on which variable?

Again, since we can not rank on both variables at the same time and some time it is easier to rank on one variable than the other, then we need to decide on which variable we should rank. We need to compare the variance of $\hat{R}_{SERSS1}$ in (4.3) and variance of $\hat{R}_{SERSS2}$ in (4.6).

Theorem 4.2 : Assume that there are L linear relations between Y_h and X_h, i.e., $|\rho_h| > 0$ and it is easy to rank on variable X. Also assume that the approximation to the variance of the ratio estimators $\hat{R}_{SERSS1}$ and $\hat{R}_{SERSS2}$ given in equations (4.8) and (4.11) respectively are valid and the bias of the estimators can be ignored, and if underlying distribution is Normal or Logistic distribution, then $Var(\hat{R}_{SERSS2}) \le Var(\hat{R}_{SERSS1})$

Proof: The proof is similar to that of Theorem 3.2. Finally in this case it is always recommended to rank on variable that will be used in the denominator of the ratio estimator if we wish to estimate the mean or total of the populaton using the ratio estimator method when the data is selected according SERSS method.

Simulation Study

Computer simulation was conducted to gain insight in the properties of the ratio estimator. Bivariate random observations were generated from a bivariate normal distribution with parameters $\mu_X, \mu_Y, \sigma_X, \sigma_Y$ and correlation coefficient ρ. Also we deviled the data into three strata and in some cases into four strata. The sampling methods described above are used to draw SERSS, SRSS and SSRS with sets of size r. We repeat this process m times to get samples of size $n = rm$. The simulation was performed with $r = 20$, 30, and 40 and with $m = 10$ for the SERSS, SRSS and SSRS data sets. The ratio of the population means was estimated from these samples. Using 2000 replications, estimates of the means, and mean square errors were computed.

The ranking was considered on either variable *Y* or *X* i.e., the ranking in one of the two variables would be perfect while the second with errors in ranking. Results of these simulations are summarized by the relative efficiencies of the estimators of the population ratio and by the bias of estimation for different values of the correlation coefficient ρ. Introduced here is only one table for efficiency when ranking on *X* and one for the bias, since other tables give the same conclusion Results of the simulation is given in Table 2 for the efficiency when ranking on variable *X*. Table 3 shows the bias of the estimators when ranking on the variable *X*. The efficiency of the ratio estimator is defined by

$$eff(R_{SSRS}, R_{SERSS}) = \frac{MSE(R_{SSRS})}{MSE(R_{SERSS})}.$$

Results

It is concluded that the highest gain in efficiency is obtained by ranking of the variable X and with large values of negative ρ. For example in Table 2, $eff(R_{SSRS}, R_{SERSS})$ when (ρ=.90, $r = 40$ & m=10) is 1.69 while when $eff(R_{SSRS}, R_{SERSS})$ (ρ=-.90, $r = 40$ & m=10) is 3.05. Also, our simulation indicates the following:

Table 2. Efficiency when ranking on variable X

W_h : .3 / .3 / .4

μ_{xh} : 2 / 3 / 4 σ_{xh} : 1 / 1 / 1

μ_{yh} : 3 / 4 / 6 σ_{yh} : 1 / 1 / 1

R= 1.45		Eff. In Combine		Eff. in Separate	
ρ	r	$\frac{SSRS}{SERSS}$	$\frac{SRSS}{SERSS}$	$\frac{SSRS}{SERSS}$	$\frac{SRSS}{SERSS}$
	20	2.1720	0.7562	2.2231	0.7499
.99	**30**	2.4724	0.6932	2.5604	0.6853
	40	2.7212	0.6762	2.7328	0.6690
	20	1.5512	0.8679	1.5715	0.8590
.90	**30**	1.7122	0.8353	1.7312	0.8354
	40	1.6920	0.8103	1.7320	0.8089
	20	1.5385	0.9894	1.5480	0.9840
.70	**30**	1.5100	0.8221	1.5243	0.8218
	40	1.4787	0.8319	1.4745	0.8280
	20	1.5079	0.9055	1.5326	0.9074
.50	**30**	1.4670	0.9231	1.4791	0.9270
	40	1.5491	0.8471	1.5630	0.8460
	20	1.5039	0.8925	1.5273	0.8921
.25	**30**	1.7163	0.8597	1.7399	0.8598
	40	1.6810	0.8471	1.6991	0.8462
	20	1.7585	0.8268	1.7780	0.8235
-.25	**30**	1.8502	0.8042	1.8654	0.8042
	40	1.9663	0.7738	1.9830	0.7751
	20	1.8968	0.7978	1.9015	0.7978
-.50	**30**	2.2485	0.7770	2.2664	0.7759
	40	2.3928	0.6844	2.4007	0.6832
	20	2.2086	1.2115	2.2175	1.2149
-.70	**30**	2.5975	1.3015	2.6208	1.3064
	40	2.4561	1.5529	2.4784	1.5572
	20	2.5302	0.7841	2.5749	0.7788
-.90	**30**	2.8748	0.7002	2.8924	0.6963
	40	3.0556	0.5546	3.0866	0.5505
	20	2.6394	0.7176	2.7023	0.7219
-.99	**30**	3.0000	0.6558	3.0266	0.6535
	40	3.1623	0.5275	3.1902	0.5269

1. When ranking on variable *X*, the efficiency will decrease with decreasing the value of ρ from 0.99 to 0.50, and start to increase as ρ decreases from 0.25 to -0.99.
2. The efficiency will increase when the even sample size increased by increasing the number of elements in each set (*r*).
3. There will be no change in the efficiency if the sample size increased by increasing the cycle size m.
4. For fix ρ, we noted that in combined case, as *r* increase the efficiency will increase, for all values of ρ positive or negative except, in some cases when (*r* =30) and ρ positive.
5. Also, for fix ρ, we note that in separate case, as *r* increase the efficiency will increase, for all values of ρ positive or negative, except in some cases when (*r* =30) and ρ positive.
6. For fix *r*, we note that in combined case the efficiency will decrease from 0.99 to 0.50, and then after this will increase from 0.25 to –0.99.
7. Also, for fix *r* and change ρ, we note that in separate case the efficiency will decrease from 0.99 to 0.50, and then after this will increase from 0.25 to -0.99.
8. We note that the efficiency in combined case less than in separate case. That is because the sample size within each stratum is small.
9. The bias will decrease when increasing the number of increases the even sample size r elements in each set.
10. The bias in combined case is less than the corresponding bias in separate case.

Application: Bilirubn Level in Jaundice Babies

Introduced is a real life example about Bilirubin level in jaundice babies who stay in neonatal intensive care. Most of birth surveys on live newborns Birth showed that jaundice is common. Jaundice in new Born can be pathological physiological which start on second day of life and it has relationship with race, method of feeding and Gestational age.

Jaundice is observed during the first week of life, and neonatal jaundice is a common problem. It is possible that the generally accepted levels are too high and may produce some high tone hearing loss.

Most of neonatal jaundice appears on second day of life. Most of normal newborn babies leave the hospital after 24 hours of life. Therefore, the primary concern will be on baby's who staying in neonatal intensive care. Physicians are interesting in the jaundice, according to its important and risk on the hearing, brain and death. We will focus on the weight and bilirubin level in blood (tsb) for the babies. The data were collected on 120 babies, who stay in neonatal intensive care, in four Jordanian hospitals (see Samawi and Al-Sagheer, 2001.) The data were divided into two strata, male stratum of size N_1=72 and female stratum of size N_2=48.

The following are the exact population values of the data

For Males it was found that: $\mu_{X1} = 2.91$, $\sigma_{X1} = 0.75$, $\mu_{Y1} = 11.97$, $\sigma_{Y1} = 5.52$ and $\rho_1 = 0.22$.

For Females it was found that: $\mu_{X2} = 2.82$, $\sigma_{X2} = 0.64$, $\mu_{Y2} = 9.97$, $\sigma_{Y2} = 4.11$ and $\rho_2 = -0.37$.

Also, for the whole data it was found that: $\mu_X = 2.87$, $\sigma_X = 0.71$, $\mu_Y = 11.18$, $\sigma_Y = 5.08$ and $\rho = 0.06$ Two strata exist (*L=2*), *m*=2 and *r*=10, which produce *n* = *r.m* = 20 $W_1 = \frac{72}{120} = 0.6$, $W_2 = \frac{48}{120} = 0.4$. For Male : $n_1 = 0.6 \times 20 = 12$; $r_1 = 6$. For Female: $n_2 = 0.4 \times 20 = 8$; $r_2 = 4$.

Table 2: Bias of the ratio estimators when ranking on variable X

$W_{h: 0..3/\ 0.3/\ 0.4}$	R=1.45	μ_{Xh} :2 / 3 / 4	μ_{Yh} :3 / 4 / 6	σ_{Xh} :1 / 1 / 1	σ_{Yh} :1 / 1 / 1		
ρ	r	Combined SRSS	Combined SSRS	Combined SERSS	Separate SRSS	Separate SSRS	Separate SERSS
0.99	20	-0.0002	0.0003	0.0001	-0.0001	0.0009	0.0003
	30	0.0000	-0.0001	0.0000	0.0001	0.0001	0.0000
	40	0.0001	0.0002	0.0000	0.0001	0.0004	0.0000
0.90	20	0.0001	0.0002	0.0003	0.0003	0.0009	0.0006
	30	0.0002	0.0007	0.0000	0.0003	0.0011	0.0001
	40	-0.0002	0.0000	0.0000	-0.0002	0.0003	0.0000
0.70	20	0.0005	0.0003	0.0007	0.0007	0.0013	0.0011
	30	0.0003	0.0000	0.0002	0.0004	0.0006	0.0004
	40	0.0004	0.0000	-0.0001	0.0005	0.0003	0.0000
0.50	20	0.0003	0.0011	-0.0005	0.0006	0.0023	-0.0001
	30	0.0007	0.0003	0.0005	0.0009	0.0011	0.0008
	40	0.0006	0.0003	0.0001	0.0007	0.0009	0.0002
0.25	20	0.0005	0.0010	0.0001	0.0010	0.0025	0.0008
	30	0.0000	-0.0006	0.0002	0.0003	0.0004	0.0005
	40	0.0002	0.0000	-0.0001	0.0003	0.0006	0.0001
-0.25	20	-0.0002	0.0007	0.0003	0.0004	0.0029	0.0012
	30	0.0005	0.0000	0.0001	0.0008	0.0013	0.0005
	40	-0.0005	-0.0003	-0.0003	-0.0003	0.0007	0.0000
-0.50	20	0.0000	-0.0011	0.0003	0.0006	0.0011	0.0013
	30	-0.0002	-0.0002	0.0000	0.000	0.0013	0.0004
	40	-0.0004	-0.0003	-0.0004	-0.0002	0.0009	0.0000
-0.70	20	0.0001	0.0006	-0.0003	0.0009	0.00031	0.0006
	30	-0.0002	0.0000	0.0000	0.0000	0.0016	0.0006
	40	0.0003	0.0007	-0.0005	0.0005	0.0020	-0.0001
-0.90	20	-0.0008	0.0022	-0.0008	0.0000	0.0053	0.0003
	30	0.0002	-0.0004	0.0001	0.0006	0.0014	0.0007
	40	0.0000	0.0021	0.0003	0.0002	0.0036	0.0008
-0.99	20	0.0015	0.0022	0.0003	0.0024	0.0053	0.0016
	30	-0.0006	0.0008	-0.0001	-0.0002	0.0028	0.0004
	40	-0.0002	0.0000	0.0000	0.0000	0.0014	0.0005

Using SERSS & SSRS

We use the method of sampling SERSS and SSRS to get the following samples. Note that the ranking was on variable *X* (weight).The SERSS which is drawn is in Table 4.

Based on the SERSS it was found that $\hat{\mu}_{X1} = 2.83$, $\hat{\mu}_{Y1} = 9.78$, $\hat{\mu}_{X2} = 3.237$, $\hat{\mu}_{Y2} = 11.74$, $\hat{\sigma}^2_{X1} = 0.286$, $\hat{\sigma}^2_{Y1} = 4.69$, $\hat{\sigma}^2_{X2} = 0.377$, $\hat{\sigma}^2_{Y2} = 12.68$.

Also, for the SSRS it was found: $\tilde{\mu}_{X1} = 2.58$, $\tilde{\mu}_{Y1} = 8.29$, $\tilde{\mu}_{X2} = 2.97$, $\tilde{\mu}_{Y2} = 11.89$, $\tilde{\sigma}^2_{X1} = 0.27$, $\tilde{\sigma}^2_{Y1} = 23.81$, $\tilde{\sigma}^2_{X2} = 0.35$, $\tilde{\sigma}^2_{Y2} = 50.41$.

Now, $\hat{V}ar\left(R_{SERSS\,2(s)}\right) = 0.071$

$$\hat{V}ar\left(R_{SERSS\,2(c)}\right) = 0.082$$

$$\hat{V}ar\left(R_{SSRS\,(s)}\right) = 0.265$$

$$\hat{V}ar\left(R_{SSRS\,(c)}\right) = 0.266$$

note that

$\hat{V}ar\left(R_{SERSS\,2(s)}\right) \leq \hat{V}ar\left(R_{SERSS\,2(c)}\right)$ and

$$\hat{V}ar\left(R_{SSRS\,(s)}\right) \leq \hat{V}ar\left(R_{SSRS(c)}\right)$$

It is clear that this just illustration of the computations only. However, still this conclusion indicates that the results in Sections 3, 4 and 5 are correct.

$$eff\,(R_{SSRS\,2(s)}, R_{SERSS\,2(s)}) = 3.73,$$

$$eff\,(R_{SSRS(c)}, R_{SERSS\,2(c)}) = 3.23$$

$$eff\,(R_{SSRS(c)}, R_{SSRS(s)}) = 1.00$$

$$eff\,(R_{SERSS\,2(c)}, R_{SERSS\,2(s)}) = 1.15\,.$$

Table 4. The drawn samples using SERSS and SSRS methods.

Cycle Number	Females		Males	
	SERSS Sample			
	tsb	Weight	tbs	Weight
	4.80	4.15	7.06	2.80
	6.90	3.00	5.60	2.75
1	7.80	3.15	5.50	3.70
	8.60	3.40	7.53	2.50
			9.50	3.60
			9.20	1.85
	12.76	2.50	23.41	3.10
	8.82	1.55	10.24	3.50
	13.94	2.85	13.18	4.50
2	14.59	2.10	14.00	3.10
			16.20	3.65
			19.50	3.80
		SSRSS		
	9.30	2.80	7.70	2.60
	5.50	3.00	6.12	3.20
1	7.80	3.15	21.29	4.15
	5.40	2.65	10.94	2.60
			9.50	3.60
			15.47	2.70
	9.24	2.60	8.71	2.45
	20.41	2.10	7.06	2.80
2	13.10	2.85	7.60	2.20
	8.82	1.55	13.60	2.50
			29.24	3.15
			5.50	3.70

References

Arnold, B. C., Balakrishnan, N., & Nagaraja, H. N. (1992). *A first course in order statistics.* NY: John Wiley & Sons.

Cochran, W. G. (1977). *Sampling techniques*. 3rd edition. NY: John Wiley & Sons.

McIntyre, G. A. (1952). A method of unbiased selective sampling, using ranked sets. *Australian Journal of Agricultural Research 3*, 385-390.

Samawi, H. M. (1996). Stratified ranked set sample. *Pakistan Journal of Statistics, 12*, 1, 9-16.

Samawi, H. M., Ahmed, M. S. & Abu-Dayyeh, W. (1996). Estimating the population mean using extreme ranked set sampling. *Biometrical Journal*, *38*, 5, 577 – 586.

Samawi, H. M. and Al-Sageer, O. A. M. (2001). On the estimation of the distribution function using extreme and median ranked set sampling. *Biometrical Journal*, *43*, 2, 29-45 .

Samawi, H. M., & Muttlak, H. A. (1996). Estimation of ratio using rank set sampling. *Biometrical Journal*, *38*, 6, 753-764 .

Takahasi, K. & Wakimoto, K. (1968). On unbiased estimates of the population mean based on the stratified sampling by means of ordering. *Annals of the Institute of Statistical Mathematics*, *20*, 1-31.