Ratio Type Estimation Using the Knowledge of the Auxiliary Variable for Ranking and Estimating

Carlos N. Bouza[1] Amer Ibrahim Al-Omari[2] Agustín Santiago[3] & Jose M. Sautto[3]

[1] Universidad de La Habana, Cuba

[2] Al al-Bayt University, Faculty of Science, Department of Mathematics, Jordan

[3] Universidad Autónoma de Guerrero, Mexico

Correspondence: Agustín Santiago, Universidad Autónoma de Guerrero, Mexico. E-mail: asantiago@uagro.mx

Abstract

In this paper, the behavior of ranked set sampling is analyzed considering the knowledge of the auxiliary variable. The suggested estimators are compared with their simple random sampling counterparts. A numerical study is developed using data from a study developed on the contamination due to burning compost from solid waste from hospitals.

Keywords: Ratio type estimators, Ranked set sampling, Simple random sampling, Gain in accuracy, Relative precision.

1. Introduction

1.1 Problem Statement

Estimation of the population parameters is widely considered in statistical inference. Good estimators are considered as those with good properties, for example, that are unbiased, efficient, and consistent. There are unbiased estimators that are good estimators due to their very small mean squared error, such as ratio estimators, which were considered in this study.

When there is not auxiliary information available, using simple random sampling is commonly the only solution in statistical inference. When we have information about an auxiliary variable that is correlated with the variable of interest, other options are possible. This information about the auxiliary variable can be used to obtain a more efficient estimator from the population mean using a ratio estimator or regression estimators.

We will denote the finite population of interest as U, with N different and identified units, which means $U = \{1, \cdots, j, \cdots, N\}$. The list that allows us to identify each one of the *population* units is called *frame*. The population units have many characteristics of interest, some of them known and some others unknown. We will represent for variable Y the population characteristic that we want to study, and we are going to name it, variable of study. The value that takes such variable over the population is unknown, but it will be given for the value that each unit j allocates to such characteristic y_j, which means, $y = \{y_1, y_2, \cdots, y_j, \cdots, y_N\}$.

Moreover, we can register the known characteristics of a variable X, that we will suppose p-dimensional, to provide of p characteristics of each population unit. Thus, for each unit j in the population, we have a vector of additional information, $x_j = (x_{j1}, \cdots, x_{jp})$. For this, the vector, $x = \{x_1, \cdots, x_N\}$ receives the name of *additional information vector*. This auxiliary information, that it is supposed to be related with the variable of interest, must be used to help to know the associated values with the variable of interest.

With this auxiliary information, we can find ratio type estimators to improve the precision of the estimation in SRS, SRSWOR or any other sampling design.   Additional improvements in the ratio estimator are also obtained through the introduction of modified ratio estimators proposed by Singh, Tailor and Tailor (2010) and Al-Omari Amer I., Jaber Kalifa and Al-Omari Ahmad (2008).

1.2 Importance of the Problem

In applications, obtaining information from an additional variable is often costly, but classification of observations according to it, is relatively easy.

We assume that the "Ranked Set Sampling" (RSS) sample design can be used to improve the accuracy of estimations, regarding the simple random sampling (SRS), and at the same time maintaining the cost or time limit on sampling. The RSS was first applied by McIntyre (McIntyre, 1952), in his study about the estimation of mean yields of pastures, and later

developed by different statisticians.

In this paper we consider that due to the relationship between X and Y, RSS can be used to estimate and to classify. Singh et al. (Singh, et al., 2010) proposed a class of ratio type estimators, which are expanded using our proposal and we believe that they are more accurate due to the reduction of the sampling error, for both, the ratio type estimator and the sampling design type RSS.

Therefore, our objective is to propose new ratio estimators for the population mean generated from an RSS design, to compare them with modified ratio estimators proposed by Singh et al. (Singh, et al., 2010) and Al-Omari et al. (Al-Omari, et al., 2008) and establish the gain in precision using the relative efficiency (*RE*) of the estimators.

1.3 Initial Definitions.

Let $\bar{X} = \frac{\sum_{i=1}^N X_i}{N}$ and $\sigma_X^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N}$ be the population mean and variance of the auxiliary variable *X*; $\bar{Y} = \frac{\sum_{i=1}^N Y_i}{N}$ and

$\sigma_Y^2 = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^2}{N}$ are the population mean and variance of the study variable *Y*; take ρ as the correlation coefficient

between *X* and *Y*. Using Simple Random Sampling with Replacement (SRSWR) for selecting a sample of size n are unbiased estimators

$$\bar{z} = \frac{\sum_{i=1}^n z_i}{n}, s_X^2 = \frac{\sum_{i=1}^n (z_i - \bar{z})^2}{n-1}, z_i = x_i, y_i$$

As we know $X_1, \dots, X_N$ and its distribution we are able to determine $X_m = Min\{X_1, \dots, X_N\}$, $X_M = \{X_1, \dots, X_N\}$ and the needed quartiles. We will consider only the first and third quartiles of the auxiliary variable *X* denoted by q_1 and q_3 These values are perfectly known before sampling. Once a sample of size m is selected using SRSWR we are able to rank them using the known values of *X* before measuring *Y*. Hence ranked set sampling (RSS) can be used as an alternative sampling design.

2. Methodology

This is a theoretical and applied research, due to the new modified ratio estimators are proposed, for which approximate variance and bias are estimated using a Taylor series expansion. Once the approximate variance is estimated, the mean squared error of each estimator is found. In order to compare the estimators, the mean squared errors (MSE) of the estimators are estimated and are compared by quotient, that is $RE = \frac{MSE(\hat{\theta}_1)}{MSE(\hat{\theta}_2)}$, where the estimator $\hat{\theta}_2$ will be more efficient than the estimator $\hat{\theta}_1$ if $RE > 1$.

2.1 The SRSWR Estimators

Singh et al. (Singh, et al., 2010) proposed the use of the transformed variables

$$u_i = \frac{x_i + X_m}{X_m + X_M}, \qquad w_i = \frac{x_i + X_M}{X_m + X_M}$$

The corresponding expectations are

$$E(u_i) = \frac{E(x_i) + X_m}{X_m + X_M} = \frac{\bar{X} + X_m}{X_m + X_M} = \bar{U},$$

$$E(w_i) = \frac{E(x_i) + X_M}{X_m + X_M} = \frac{\bar{X} + X_M}{X_m + X_M} = \bar{W}$$

Hence $\bar{u} = \frac{\bar{x} + X_m}{X_m + X_M}$ and $\bar{w} = \frac{\bar{x} + X_M}{X_m + X_M}$ are unbiased estimators

The ratio type estimators proposed by them are

$$\bar{y}_u = \bar{y}\left(\frac{\sum_{i=1}^N U_i/N}{\sum_{i=1}^n u_i/n}\right) = \bar{y}\left(\frac{\bar{U}}{\bar{u}}\right) = r_u \bar{U}$$

$$\bar{y}_w = \bar{y}\left(\frac{\sum_{i=1}^N W_i/N}{\sum_{i=1}^n w_i/n}\right) = \bar{y}\left(\frac{\overline{W}}{\overline{w}}\right) = r_w \overline{W}$$

Al-Omari et al. (2008) suggested under SRSWR

$$\bar{y}_{q_h} = \bar{y}\left(\frac{\sum_{i=1}^N \frac{X_i}{N} + q_h}{\sum_{i=1}^n \frac{x_i}{n} + q_h}\right), h = 1,3$$

Using Taylor series approximations the approximate biases were derived. Take $C_Z = \frac{\sigma_Z}{\bar{Z}}, Z = X, Y$ The next theorem gives an approximation to the biases.

Theorem: For the first degree of approximation

$$B(\bar{y}_u) \cong \frac{R}{n}\left(\frac{\sigma_X^2}{\overline{X} + X_m}\right)\left(\frac{\overline{X}}{\overline{X} + X_m} - \rho\frac{C_Y}{C_X}\right)$$

$$B(\bar{y}_w) \cong \frac{R}{n}\left(\frac{\sigma_X^2}{\overline{X} + X_M}\right)\left(\frac{\overline{X}}{\overline{X} + X_M} - \rho\frac{C_Y}{C_X}\right)$$

$$B(\bar{y}_{q_h}) \cong 0, \qquad h = 1,3$$

Proof:

See Singh et al (Singh, et al., 2010) and Al-Omari et al (Al-Omari, et al., 2008). ▢

The approximate Mean Squared Errors (MSE) are given by

Theorem: For the first degree of approximation

$$M_u = MSE(\bar{y}_u) \cong \frac{\overline{Y}^2}{n}\left(C_Y^2 + \frac{C_X^2}{C_m^2} - 2\rho\frac{C_X C_Y}{C_m}\right), C_m = 1 + \frac{X_m}{\overline{X}}$$

$$M_w = MSE(\bar{y}_w) \cong \frac{\overline{Y}^2}{n}\left(C_Y^2 + \frac{C_X^2}{C_M} - 2\rho\frac{C_Y}{C_X}\right), C_M = 1 + \frac{X_M}{\overline{X}}$$

Proof:

See Singh et al (2010) and Bouza- Al-Olomari (2011). ▢

The transformation

$$a_i = X_M x_i + X_m^2$$

leads to

$$\bar{y}_a = \bar{y}\left(\frac{\sum_{i=1}^N A_i/N}{\sum_{i=1}^n a_i/n}\right) = \bar{y}\left(\frac{\overline{A}}{\bar{a}}\right) = r_a \overline{A}$$

$$\bar{a} = X_M \bar{x} + X_m^2, \quad \overline{A} = X_M \overline{X} + X_m^2$$

Take

$$\bar{z} = \overline{Z}(1 + e_z),$$

$$E(e_z) = 0, E(e_z^2) = \frac{C_Z^2}{n}, Z = X, Y, E(e_X e_Y) = \frac{\rho C_X C_Y}{n},$$

As

$$\bar{y}_a = \overline{Y}(1 + e_y)\left(\frac{\sum_{i=1}^N \frac{A_i}{N}}{X_M \overline{X}(1 + e_x) + X_m^2}\right) = \overline{Y}(1 + e_y)(1 + \gamma(X)e_x)^{-1}, \gamma(a) = \frac{X_M \overline{X}}{X_M \overline{X} + X_m^2}$$

Assuming that $|\gamma(X)e_x| < 1$ is valid to use the approximation

$$\bar{y}_a \cong \overline{Y}(1 + e_y)\left(\frac{\sum_{i=1}^N \frac{A_i}{N}}{X_M \overline{X}(1 + e_x) + X_m^2}\right) = \overline{Y}(1 + e_y - \gamma(a)e_x - \gamma(a)e_x e_y +, \gamma(a)^2 e_x^2)$$

Then is derived that the bias is given, approximately, by

$$B(\bar{y}_a) \cong \frac{\bar{Y}\left(C_X^2 \gamma(a)\right)\left(\gamma(a) - \rho \frac{C_Y}{C_X}\right)}{n}$$

and the MSE by

$$M_a = MSE(\bar{y}_a) \cong \frac{\bar{Y}^2\left(C_Y^2 + C_X^2 \gamma(a)\right)\left(\gamma(a) - 2\rho \frac{C_Y}{C_X}\right)}{n}$$

Another transformation is based on

$$b_i = (X_M - X_m)x_i + X_m^2$$

and the ratio type estimator derived using it is

$$\bar{y}_b = \bar{y}\left(\frac{\sum_{i=1}^N B_i / N}{\sum_{i=1}^n b_i / n}\right) = \bar{y}\left(\frac{\bar{B}}{\bar{b}}\right) = r_b \bar{B}$$

Using $\gamma(b) = \frac{(X_M - X_m)\bar{X}}{\sum_{i=1}^N B_i / N}$, its bias and MSE are approximately,

$$B(\bar{y}_b) \cong \frac{\bar{Y}\left(C_X^2 \gamma(b)\right)\left(\gamma(b) - \rho \frac{C_Y}{C_X}\right)}{n}$$

and

$$M_b = MSE(\bar{y}_b) \cong \frac{\bar{Y}^2\left(C_Y^2 + C_X^2 \gamma(b)\right)\left(\gamma(b) - 2\rho \frac{C_Y}{C_X}\right)}{n}$$

2.2 RSS Estimators

The ranked set sampling (RSS) was first suggested by McIntyre (McIntyre, 1952) for estimating the population mean of forage yields. He claimed that it was a more efficient method than the commonly used simple random sampling (SRS). The ranked set sampling method can be described as follows:

The theoretical frame that permits use of the RSS model is based on the hypothesis

i We wish to enumerate the measurable variable Y.

ii The units can be ordered linearly without ties.

iii Any sample $s \subset U$ of size m can be enumerated.

iiii To identify a unit, order the units in s. To enumerate them is less costly than to evaluate $\{Y_i, i \in s\}$ or to order U.

In survey sampling settings it is logic ranking the units based on the values of an auxiliary variable correlated with the variable of interest. The basic RSS procedure is the following:

Step 1: Randomly select m^2 units from the target population. These units are randomly allocated into m sets, each of size m.

Step2: The m units of each set are ranked visually or by any inexpensive method free of cost, say X, with respect to the variable of interest Y.

Step2: From the first set of m units, the smallest ranked unit is measured; from the second set of m units the second smallest ranked unit is measured.

Step 3: Continue until the mth smallest unit (the largest) is measured from the last set.

Step 4: Repeat the whole process $r(i)$ times (cycles)

Step 5: Evaluate the corresponding units.

We can denote it a follows

$$\left.\begin{array}{c} X_{i1} \\ X_{i2} \\ \vdots \\ X_{im} \end{array}\right\}_r \sim > X_{(ii)r} \text{ and mesure } Y_{(ii)r,} \ r = 1, \dots, r(i); \ i = 1, \dots, m$$

Let $Y_1, \cdots, Y_m$, be a sample selected using SRSWR from probability density function $f(y)$, with mean μ_Y and variance σ_Y^2. Considering the selection of m independent samples selected using a SRSWR design, each of size m, we have $Y_{11}, \cdots, Y_{1m}, Y_{21}, \cdots, Y_{2m}, \cdots, Y_{m1}, \cdots, Y_{mm}$. Let $Y_{i(1m)}, \cdots, Y_{i(mm)}$, be the order statistics of the sample $Y_{1i}, \cdots, Y_{1m}, \cdots, Y_{im}$ Y1i, for $(i = 1, \cdots, m)$. The probability density function (pdf) and cumulative distribution function (cdf), mean and variance of the jth order statistics $Y_{(jm)}$, are given by

$$f_{(jm)}(y) = \frac{m!}{(j-1)!\,(m-j)!} \left(F(y)\right)^{j-1} \left(1 - F(y)\right)^{m-j}$$

$$F_{(jm)}(y) = \frac{m!}{(j-1)!\,(m-j)!} \int_0^{F(x)} v^{j-1}(1-v)^{m-j} dv$$

$$\mu_{Y_{(jm)}} = \int_{-\infty}^{\infty} y f_{(jm)}(y) dy, \quad \sigma_{Y_{(jm)}}^2 = \int_{-\infty}^{\infty} (y - \mu_{Y_{(jm)}})^2 f_{(jm)}(y) dy,$$

respectively, see David and Nagaraja (2003).

Takahasi and Wakimoto (1968) provided the mathematical theory of RSS and showed that

$$f(y) = \frac{\sum_{j=1}^{m} f_{(jm)}(y)}{m}$$

$$\mu_Y = \frac{\sum_{j=1}^{m} \mu_{Y_{(jm)}}(y)}{m},$$

and

$$V(Y_{(jm)} = \sigma_Y^2 - \Delta_{Y_{(jm)}}^2, \Delta_{Y_{(jm)}}^2 = \left(\mu_{Y_{(jm)}} - \mu_Y\right)^2, \qquad j = 1, \dots, m$$

Without losing in generality we will drop the value m of the sample size in the notation in the sequel when it provides no further information.

Note that if $r = 1$ we observe only a RSS of size $m = n$.

We will consider generically $\bar{z}_{RSS} = \frac{\sum_{t=1}^{r} \sum_{i=1}^{m} z_{(i)t}}{rm}, rm = n$ then a RSS counterpart of $\bar{y}_u$ is

$$\bar{y}_{uRSS} = \bar{y}_{RSS} \left(\frac{\overline{U}}{\overline{u}_{RSS}}\right) = r_{uRSS} \overline{U},$$

Take

$$\bar{z}_{RSS} = \bar{Z}(1 + e_{zRSS}), e_{zRSS} = \frac{\bar{z}_{RSS} - \bar{Z}}{\bar{Z}}$$

It is clear that

$$E(e_{zRSS}) = 0, E(e_{zRSS}^2) = \frac{V(\bar{z}_{RSS})}{\bar{Z}^2},$$

$$Cov(\bar{y}_{RSS}\bar{x}_{RSS}) = \rho \sqrt{\left(\frac{V(\bar{y}_{RSS})}{\bar{Y}} \times \frac{V(\bar{x}_{RSS})}{\bar{X}}\right)} = \rho C_{Y_{RSS}} C_{X_{RSS}},$$

We can write

$$\bar{z}_{RSS} = \bar{Z}\left(1 + \frac{\bar{z}_{RSS} - \bar{Z}}{\bar{Z}}\right)$$

$$\bar{y}_{uRSS} = \bar{Y}\left(1 + e_{yRss}\right)\left(\frac{\overline{U}}{\overline{u}_{RSS}}\right) = \frac{\bar{Y}\left(1 + e_{yRss}\right)}{1 + \tau(u) e_{xRSS}}, \tau(u) = \frac{\bar{X}}{\bar{X} + X_m}$$

We will consider the terms of order larger than 2 to be negligible in the needed Taylor series developments.

If $|\tau(u) e_{xRSS}| < 1$ we can use a Taylor Series development $(1 + \tau(u) e_{xRSS})^{-1}$ and:

$$E(\bar{y}_{uRSS}) \cong \bar{Y}E(1 + e_{yRss})(1 - \tau(u)e_{xRSS}) = \bar{Y} + \bar{Y}\left[-\tau(u)E(e_{yRSS}e_{xRSS}) + \tau^2(u)E(e_{xRSS}^2)\right])$$

$$= \bar{Y} + \bar{Y}\left(-\tau(u)\rho\sqrt{\left(\frac{V(\bar{y}_{RSS})}{\bar{Y}^2} \times \frac{V(\bar{x}_{RSS})}{\bar{X}^2}\right)} + \frac{\tau^2(u)V(\bar{x}_{RSS})}{\bar{X}^2}\right)$$

Hence

$$B(\bar{y}_{uRSS}) \cong \bar{Y}\left(\tau^2(u)\frac{\frac{\sigma_X^2}{n} - \frac{\sum_{j=1}^m \Delta_{X_{(j)}}^2}{mn}}{\bar{X}^2} - \frac{\tau(u)\rho\sqrt{\left(\frac{\sigma_X^2}{n} - \frac{\sum_{j=1}^m \Delta_{X_{(j)}}^2}{mn}\right)\left(\frac{\sigma_Y^2}{n} - \frac{\sum_{j=1}^m \Delta_{Y_{(j)}}^2}{mn}\right)}}{\bar{Y}\bar{X}}\right)$$

$$= \frac{R}{n}\left(\frac{\bar{X}\sigma_X^2}{(\bar{X} + X_m)^2} - \frac{\bar{X}\sum_{j=1}^m \Delta_{X_{(j)}}^2}{m(\bar{X} + X_m)^2}\right) - \frac{R\bar{X}\rho}{n(\bar{X} + X_m)}\left(\sqrt{\left(\sigma_X^2 - \frac{\sum_{j=1}^m \Delta_{X_{(j)}}^2}{m}\right)}\sqrt{\left(C_Y^2 - \frac{\sum_{j=1}^m \Delta_{Y_{(j)}}^2}{m\bar{Y}^2}\right)}\right)$$

Let us consider the MSE. Now

$$M_{uRSS} = MSE(\bar{y}_{uRSS}) = E(\bar{y}_{uRSS} - \bar{Y})^2 \cong \bar{Y}^2 E(e_{yRSS}^2) + \tau^2(u)(e_{xRSS}^2) - 2\tau(u)[e_{yRSS}e_{xRSS}]) = \frac{\sigma_Y^2}{n} - \frac{\sum_{j=1}^m \Delta_{Y_{(j)}}^2}{mn} +$$

$$\frac{R^2}{n}\tau^2(u)\left(\sigma_X^2 - \frac{\sum_{j=1}^m \Delta_{X_{(j)}}^2}{m}\right)$$

$$-2\frac{(R\tau(u)\rho)}{n}\left(\sqrt{\left(\sigma_X^2 - \frac{\sum_{j=1}^m \Delta_{X_{(j)}}^2}{m}\right)\left(\sigma_Y^2 - \frac{\sum_{j=1}^m \Delta_{Y_{(j)}}^2}{m}\right)}\right)$$

These results support the following proposition

Proposition. Consider the use of an RSS design of size $n = mr$, $\tau(u) = \frac{\bar{X}}{\bar{X} + X_m}$ and the estimator

$$\bar{y}_{uRSS} = \bar{y}_{RSS}\left(\frac{\bar{U}}{\bar{u}_{RSS}}\right) = r_{uRSS}\bar{U},$$

Then

$$B(\bar{y}_{uRSS}) \cong \frac{R}{n}\left(\frac{\bar{X}\sigma_X^2}{(\bar{X} + X_m)^2} - \frac{\bar{X}\sum_{j=1}^m \Delta_{X_{(j)}}^2}{m(\bar{X} + X_m)^2}\right) - \frac{R\bar{X}\rho}{n(\bar{X} + X_m)}\left(\sqrt{\left(\sigma_X^2 - \frac{\sum_{j=1}^m \Delta_{X_{(j)}}^2}{m}\right)}\sqrt{\left(C_Y^2 - \frac{\sum_{j=1}^m \Delta_{Y_{(j)}}^2}{m\bar{Y}^2}\right)}\right)$$

$$MSE(\bar{y}_{uRSS}) \cong \frac{\sigma_Y^2}{n} - \frac{\sum_{j=1}^m \Delta_{Y_{(j)}}^2}{mn} + \frac{R^2}{n}\tau^2(u)\left(\sigma_X^2 - \frac{\sum_{j=1}^m \Delta_{X_{(j)}}^2}{m}\right) - 2\frac{(R\tau(u)\rho)}{n}\left(\sqrt{\left(\sigma_X^2 - \frac{\sum_{j=1}^m \Delta_{X_{(j)}}^2}{m}\right)\left(\sigma_Y^2 - \frac{\sum_{j=1}^m \Delta_{Y_{(j)}}^2}{m}\right)}\right)\square$$

Remark 1. Rewriting

$$B(\bar{y}_u) \cong \frac{R}{n}\left(\frac{\sigma_X^2}{\bar{X} + X_m}\right)\left(\frac{\bar{X}}{\bar{X} + X_m}\right) - \frac{R}{n}\bar{X}\rho\left(\frac{\sigma_X}{\bar{X} + X_m}\right)C_Y$$

is evidenced that

$$B(\bar{y}_{uRSS}) \leq B(\bar{y}_u)$$

Remark 2. As

$$\frac{\bar{Y}^2}{n}\left(C_Y^2 + \frac{C_X^2}{C_m^2}\right) \geq \frac{\sigma_Y^2}{n} - \frac{\sum_{j=1}^m \Delta_{Y_{(j)}}^2}{mn} + \frac{\bar{Y}^2}{n(\bar{X} + X_m)^2}\left(\sigma_X^2 - \frac{\sum_{j=1}^m \Delta_{X_{(j)}}^2}{m}\right)$$

the two first terms of

$$MSE(\bar{y}_u) \cong \frac{\bar{Y}^2}{n}\left(C_Y^2 + \frac{C_X^2}{C_m{}^2} - 2\rho \frac{C_X C_Y}{C_m}\right)$$

are larger than the corresponding terms of the MSE of $\bar{y}_{uRSS}$. The last term can be rewritten, as $C_m = 1 + \frac{X_m}{\bar{X}} = \frac{\bar{X} + X_m}{\bar{X}}$, obtaining that $\frac{R}{n}\left(2\rho \frac{\bar{X}\sigma_X \sigma_Y}{\bar{X}+X_m}\right)$. Hence

$$2\frac{(R\tau(u)\rho)}{n}\left(\sqrt{\left(\sigma_X^2 - \frac{\sum_{j=1}^m \Delta_{X_{(j)}}^2}{m}\right)\left(\sigma_Y^2 - \frac{\sum_{j=1}^m \Delta_{Y_{(j)}}^2}{m}\right)}\right)$$

$$= 2\frac{(R\tau(u)\rho)}{n}\frac{\bar{X}\sigma_X\sigma_Y}{\bar{X}+X_m}\left(\sqrt{1 - \frac{\sum_{j=1}^m \Delta_{Y_{(j)}}^2}{m\sigma_X^2} - \frac{\sum_{j=1}^m \Delta_{X_{(j)}}^2}{m\sigma_Y^2} + \frac{\sum_{j=1}^m \Delta_{Y_{(j)}}^2}{m\sigma_X^2} \times \frac{\sum_{j=1}^m \Delta_{X_{(j)}}^2}{m\sigma_Y^2}}\right)$$

Therefore, a sufficient condition for having the last term of the bias of $\bar{y}_u$ smaller than the above equation is

$$\frac{\sum_{j=1}^m \Delta_{Y_{(j)}}^2}{m\sigma_X^2} \times \frac{\sum_{j=1}^m \Delta_{X_{(j)}}^2}{m\sigma_Y^2}\left(1 - \frac{\sigma_Y^2}{\sigma_X^2 \sum_{j=1}^m \Delta_{X_{(j)}}^2} - \frac{\sigma_X^2}{\sigma_Y^2 \sum_{j=1}^m \Delta_{Y_{(j)}}^2}\right) > 0$$

In such cases we have that $\bar{y}_{uRSS}$ is more precise than $\bar{y}_u$.

Considering $\bar{y}_w = \bar{y}\left(\frac{\bar{W}}{\bar{w}}\right) = r_w \bar{W}$ the RSS counterpart is

$$\bar{y}_{wRSS} = \bar{y}_{RSS}\left(\frac{\bar{W}}{\bar{w}_{RSS}}\right) = r_{wRSS}\bar{W}$$

Developing a similar analysis we have

Proposition. Consider the use of an RSS design of size $n = mr$, $\tau(w) = \frac{\bar{X}}{\bar{X}+X_M}$ and the estimator $\bar{y}_{wRSS}$. Then

$$B(\bar{y}_{wRSS}) \cong \frac{R}{n}\left(\frac{\bar{X}\sigma_X^2}{(\bar{X}+X_M)^2} - \frac{\bar{X}\sum_{j=1}^m \Delta_{X_{(j)}}^2}{m(\bar{X}+X_M)^2}\right) - \frac{R\bar{X}\rho}{n(\bar{X}+X_M)}\left(\sqrt{\left(\sigma_X^2 - \frac{\sum_{j=1}^m \Delta_{X_{(j)}}^2}{m}\right)}\sqrt{\left(C_Y^2 - \frac{\sum_{j=1}^m \Delta_{Y_{(j)}}^2}{m\bar{Y}^2}\right)}\right)$$

$$M_{wRSS} = MSE(\bar{y}_{wRSS}) \cong$$

$$\frac{\sigma_Y^2}{n} - \frac{\sum_{j=1}^m \Delta_{Y_{(j)}}^2}{mn} + \frac{R^2}{n}\tau^2(w)\left(\sigma_X^2 - \frac{\sum_{j=1}^m \Delta_{X_{(j)}}^2}{m}\right) - 2\frac{(R\tau(w)\rho)}{n}\left(\sqrt{\left(\sigma_X^2 - \frac{\sum_{j=1}^m \Delta_{X_{(j)}}^2}{m}\right)\left(\sigma_Y^2 - \frac{\sum_{j=1}^m \Delta_{Y_{(j)}}^2}{m}\right)}\right)\square$$

Similar remarks can be fixed easily but it is worth noting that using $\bar{y}_{wRSS}$ is better than $\bar{y}_{uRSS}$ alternative as $X_M > X_m$. A recommendation it to use an auxiliary variable with a large range $X_M - X_m$.

Let us look for the RSS version of $\bar{y}_a$. We propose

Proposition. Consider the use of an RSS design of size $n = mr$, and the estimator

$$\bar{y}_{aRSS} = \bar{y}_{RSS}\left(\frac{\bar{A}}{\bar{a}_{RSS}}\right) = r_{aRSS}\bar{A}$$

Its bias and MSE are

$$B(\bar{y}_{aRSS}) \cong \frac{R}{n}\left(\frac{\gamma(a)^2}{\bar{X}}\left(\sigma_X^2 - \frac{\sum_{j=1}^m \Delta_{X_{(j)}}^2}{m}\right) - \frac{\gamma(a)}{\bar{Y}}\rho\sqrt{\left(\frac{\sigma_X^2}{n} - \frac{\sum_{j=1}^m \Delta_{X_{(j)}}^2}{mn}\right)\left(\frac{\sigma_Y^2}{n} - \frac{\sum_{j=1}^m \Delta_{Y_{(j)}}^2}{mn}\right)}\right)$$

and

Proof:

Take

$$\bar{y}_{aRSS} \cong \bar{Y}(1 + e_{yRSS})\left(\frac{\sum_{i=1}^{N}\frac{A_i}{N}}{X_M\bar{X}(1 + e_{xRSS}) + X_m^2}\right) = \bar{Y}(1 + e_{yRSS} - \gamma(a)e_{xRSS} - \gamma(a)e_{xRSS}e_{yRSS} + \gamma(a)^2 e_{xRSS}^2$$

$$\gamma(a) = \frac{X_m\bar{X}}{\bar{A}}$$

Then

$$B(\bar{y}_{aRSS}) \cong \bar{Y}\big(E(e_{yRss}) - \gamma(a)E(e_{yRSS}e_{xRSS}) + \gamma(a)^2 E(e_{xRSS}^2)\big)$$

$$= \bar{Y}\left(-\gamma(a)E\left(\frac{\bar{y}_{RSS} - \bar{Y}}{\bar{Y}} \times \frac{\bar{x}_{RSS} - \bar{X}}{\bar{X}}\right) + \gamma(a)^2 V(e_{xRSS}^2)\right)$$

$$= \bar{Y}\left(-\gamma(a)E\left(\frac{Cov(\bar{y}_{RSS}, \bar{x}_{RSS})}{\bar{Y}\bar{X}}\right) + \gamma(a)^2 V(e_{xRSS}^2)\right)$$

$$= \bar{Y}\left[\left(\frac{\gamma(a)^2}{\bar{X}^2}\right)\left(\frac{\sigma_X^2}{n} - \frac{\sum_{j=1}^{m}\Delta_{X_{(j)}}^2}{mn}\right) - \frac{\gamma(a)}{\bar{Y}\bar{X}}\rho\sqrt{\left(\frac{\sigma_Y^2}{n} - \frac{\sum_{j=1}^{m}\Delta_{Y_{(j)}}^2}{mn}\right)\left(\frac{\sigma_X^2}{n} - \frac{\sum_{j=1}^{m}\Delta_{X_{(j)}}^2}{mn}\right)}\right]$$

$$, \qquad \Delta_{Z_{(j)}}^2 = \left(\mu_{Y_{(j)}} - \mu_Y\right)^2, j = 1, \ldots, m$$

Therefore

$$B(\bar{y}_{aRSS}) \cong \frac{R}{n}\left(\frac{\gamma(a)^2}{\bar{X}}\left(\sigma_X^2 - \frac{\sum_{j=1}^{m}\Delta_{X_{(j)}}^2}{m}\right) - \frac{\gamma(a)}{\bar{Y}}\rho\sqrt{\left(\frac{\sigma_X^2}{n} - \frac{\sum_{j=1}^{m}\Delta_{X_{(j)}}^2}{mn}\right)\left(\frac{\sigma_Y^2}{n} - \frac{\sum_{j=1}^{m}\Delta_{Y_{(j)}}^2}{mn}\right)}\right)$$

The MSE is obtained by calculating

$$M_{aRSS} = MSE(\bar{y}_{aRSS}) = E(\bar{y}_{aRSS} - \bar{Y})^2 \cong \bar{Y}^2\big(E(e_{yRSS}^2) + \gamma(a)^2 E(e_{xRSS}^2) - 2\gamma(a)E(e_{yRSS}e_{xRSS})\big) = \bar{Y}^2\left[\left(\frac{\sigma_Y^2}{n\bar{Y}^2} - \right.\right.$$

$$\left.\left.\frac{\sum_{j=1}^{m}\Delta_{Y_{(j)}}^2}{mn\bar{Y}^2}\right) + \gamma(a)^2\left(\frac{\sigma_X^2}{n\bar{X}^2} - \frac{\sum_{j=1}^{m}\Delta_{X_{(j)}}^2}{mn\bar{X}^2}\right) - 2\gamma(a)\rho\sqrt{\left(\frac{\sigma_X^2}{n\bar{X}^2} - \frac{\sum_{j=1}^{m}\Delta_{X_{(j)}}^2}{mn\bar{X}^2}\right)^{-1}\left(\frac{\sigma_Y^2}{n\bar{Y}^2} - \frac{\sum_{j=1}^{m}\Delta_{Y_{(j)}}^2}{mn\bar{Y}^2}\right)}\right]. \square$$

We can rewrite the bias of the SRSWR-based estimator as

$$B(\bar{y}_a) \cong \frac{R}{n}\left(\frac{\sigma_X^2}{\bar{X}}\right)\gamma(a)^2 - \frac{\bar{Y}\sigma_X^2\gamma(a)}{n\bar{X}}\rho\frac{C_Y}{C_X} = \frac{R}{n}\left(\frac{\sigma_X^2}{\bar{X}}\right)\gamma(a)^2 - \rho\gamma(a)\frac{\sigma_Y\sigma_X}{n}$$

Note that

$$\left(\sigma_X^2 - \frac{\sum_{j=1}^{m}\Delta_{X_{(j)}}^2}{m}\right) \le \sigma_X^2$$

$$\sqrt{\left(\frac{\sigma_X^2}{n} - \frac{\sum_{j=1}^{m}\Delta_{X_{(j)}}^2}{mn}\right)\left(\frac{\sigma_Y^2}{n} - \frac{\sum_{j=1}^{m}\Delta_{Y_{(j)}}^2}{mn}\right)} \le \frac{\sigma_Y\sigma_X}{n}$$

Hence the proposed RSS counterpart always has a smaller bias.

Note that

$$MSE(\bar{y}_a) \cong \frac{\bar{Y}^2 \left(C_Y^2 + C_X^2 \gamma^2(a) - 2\rho\gamma(a)\frac{C_Y}{C_X} \right)}{n}$$

The first two terms of $MSE(\bar{y}_{aRSS})$ are smaller than the corresponding ones of $MSE(\bar{y}_a)$. Hence is sufficient

accepting that $\frac{C_Y}{C_X} \cong \sqrt{\left(\frac{\sigma_X^2}{n\bar{X}^2} - \frac{\sum_{j=1}^m \Delta_{X(j)}^2}{mn\bar{X}^2} \right)^{-1} \left(\frac{\sigma_Y^2}{n\bar{Y}^2} - \frac{\sum_{j=1}^m \Delta_{Y(j)}^2}{mn\bar{Y}^2} \right)}$ for ensuring that the RSS counterpart is more accurate

that the SRSWR estimator.

Let us look for the RSS version of $\bar{y}_b$. A natural representation of the RSS counterpart is

$$\bar{y}_{bRSS} = \bar{y}_{RSS} \left(\frac{\bar{B}}{\bar{b}_{RSS}} \right) = r_{bRSS}\bar{B}$$

Again using the Taylor Series approximation we have

$$\bar{y}_{bRSS} \cong \bar{Y}(1 + e_{yRSS}) \left(\frac{\sum_{i=1}^N \frac{B_i}{N}}{X_M \bar{X}(1 + e_{xRSS}) + X_m^2} \right) = \bar{Y}(1 + e_{yRSS} - \gamma(b)e_{xRSS} - \gamma(b)e_{xRSS}e_{yRSS} + \gamma(b)^2 e_{xRSS}^2)$$

The bias and MSE are then approximated, $\gamma(b) = \frac{(X_M - X_m)\bar{X}}{\sum_{i=1}^N B_i/N}$

$$B(\bar{y}_{bRSS}) \cong \frac{1}{n} \left(\left(\frac{R\gamma(b)}{\bar{X}} \left(\sigma_X^2 - \frac{\sum_{j=1}^m \Delta_{X(j)}^2}{m} \right) \right) \left(\gamma(b) - \frac{\rho}{R} \sqrt{\left(\sigma_X^2 - \frac{\sum_{j=1}^m \Delta_{X(j)}^2}{m} \right)^{-1} \left(\sigma_Y^2 - \frac{\sum_{j=1}^m \Delta_{Y(j)}^2}{m} \right)} \right) \right)$$

and

$$M_{bRSS} = MSE(\bar{y}_{bRSS}) \cong \frac{\left(\sigma_Y^2 - \frac{\sum_{j=1}^m \Delta_{Y(j)}^2}{m} + R^2\sigma_X^2 - R^2 \frac{\sum_{j=1}^m \Delta_{X(j)}^2}{m}\gamma(b) \right) \left(\gamma(b) - \frac{2\rho}{R} \sqrt{\frac{\sigma_Y^2 - \frac{\sum_{j=1}^m \Delta_{Y(j)}^2}{m}}{\sigma_X^2 - \frac{\sum_{j=1}^m \Delta_{X(j)}^2}{m}}} \right)}{n}$$

We have proved the proposition.

Proposition. Consider the use of an RSS design of size $n = mr$, and the estimator
$$\bar{y}_{bRSS} = \bar{y}_{RSS} \left(\frac{\bar{B}}{\bar{b}_{RSS}} \right) = r_{bRSS}\bar{B}$$

Has as bias and MSE

$$B(\bar{y}_{bRSS}) \cong \frac{1}{n} \left(\left(\frac{R\gamma(b)}{\bar{X}} \left(\sigma_X^2 - \frac{\sum_{j=1}^m \Delta_{X(j)}^2}{m} \right) \right) \left(\gamma(b) - \frac{\rho}{R} \sqrt{\left(\sigma_X^2 - \frac{\sum_{j=1}^m \Delta_{X(j)}^2}{m} \right)^{-1} \left(\sigma_Y^2 - \frac{\sum_{j=1}^m \Delta_{Y(j)}^2}{m} \right)} \right) \right)$$

and

$$MSE(\bar{y}_{bRSS}) \cong \frac{\left(\sigma_Y^2 - \frac{\sum_{j=1}^m \Delta_{Y(j)}^2}{m} + R^2\sigma_X^2 - R^2 \frac{\sum_{j=1}^m \Delta_{X(j)}^2}{m}\gamma(b) \right) \left(\gamma(b) - \frac{2\rho}{R} \sqrt{\frac{\sigma_Y^2 - \frac{\sum_{j=1}^m \Delta_{Y(j)}^2}{m}}{\sigma_X^2 - \frac{\sum_{j=1}^m \Delta_{X(j)}^2}{m}}} \right)}{n}.$$

Under the hypothesis used for being sure that $\bar{y}_{aRSS}$ is more accurate than $\bar{y}_a$ the same result holds for $\bar{y}_{bRSS}$ with respect to $\bar{y}_b$.

3. A Numerical Study

We evaluated the behavior of the analyzed estimators in terms of their MSE's derived.   We used data coming from a study of leaching of elements from solid waste compost. The grab samples were prepared from multiple grab samples, using coning and quartering methods. The compost was collected from hospitals.   The particles were mechanically separated and passed through a fine.   Each batch, send for burning, was evaluated in terms of its estimated toxicity. This estimation is made from a sample analyzed by a laboratory.   A qualification in the range 0-100 was given to each batch. This qualification was used for ranking. For the experiment a sensor was placed in the chimney for measuring the content in the smoke of plumb, magnesium, cadmium and the rest was classified as "other contaminants".   The measurement was made for each batch introduced in the furnaces. The study was developed during six months. In the period 1678 batches were evaluated.

With the obtained data the ratios were computed $F(\gamma, \vartheta) = \frac{M_\gamma}{M_\vartheta}$, $\gamma = u, w, a, b$; $\vartheta = uRSS, wRSS, aRSS, bRSS$.

When $F(\gamma, \vartheta) > 1$ is preferred the sampling strategy with mean squared error equal to M_γ. The results are reported in tables 1-4.   In all the cases the RSS estimators were more efficient. The largest gains were obtained by $\bar{y}_{uRSS}$ when estimating the contamination in the case of plumb, $\bar{y}_{wRSS}$ provided the arger gains in accuracy for magenesium and $\bar{y}_{bRSS}$ for cadmium and for the rest of the contaminates

Table 1. Relative efficiency of the estimators for plumb

	Mu	MW	Ma	Mb
MuRSS	1,532	1,289	1,427	1,006
MwRSS	**2,151**	1,620	1,886	1,215
MaRSS	**2,329**	1,725	2,843	1,210
MbRSS	1,596	1,390	1,687	1,191

Table 2. Relative efficiency of the estimators for magnesium

	Mu	MW	Ma	Mb
MuRSS	1,857	**2,295**	1,197	1,009
MwRSS	**2,211**	**2,594**	1,566	1,088
MaRSS	**2,471**	**2,672**	1,671	1,106
MbRSS	**2,963**	**3,059**	1,936	1,228

Table 3. Relative efficiency of the estimators for cadmium

	Mu	MW	Ma	Mb
MuRSS	1,075	1,097	**2,142**	1,157
MwRSS	1,243	1,111	**2,390**	1,405
MaRSS	1,479	1,238	**2,561**	1,600
MbRSS	1,442	1,206	**2,449**	1,499

Table 4. Relative efficiency of the estimators for other contaminants

	Mu	MW	Ma	Mb
MuRSS	1,163	1,673	1,988	1,775
MwRSS	1,884	**2,117**	**2,996**	1,882
MaRSS	1,376	1,732	**3,006**	1,984
MbRSS	**2,552**	**2,769**	**3,948**	**2,716**

4. Conclusion

According to the obtained results in the numerical study, all ratio estimators constructed from an RSS design are at least as efficient as those proposed by Singh et al. (Singh, et al., 2010) and Al-Omari et al. (Al-Omari, et al., 2008), so the recommendation is that, the estimators proposed by us are used in the future, because, they do not represent an increase in cost and time of the sampling in the estimation phase and its mean squared errors are lower than those obtained by direct estimation or by modified ratio type estimation discussed in the present study for simple random sampling. We consider that, due to the relationship between X and Y the RSS design can be used to estimate as well as to classify. The ratio type estimators are an extension to those proposed by Singh et al. (Singh, et al., 2010), such extension resulted to be more accurate due to the reduction of the sampling error, for both; the use of ratio type estimators and for the RSS sampling design.

References

Al,-O. A. A.-O. (2008). Modified ratio-type estimators of the mean using extreme ranked set sampling. *Journal of Mathematics and Statistics, 4*, 150-155.

Bai, Z. D. (2003). On the theory of ranked set sampling and its ramifications. *Journal of Statistical Planning and Inference,* (109), 81-99. https://doi.org/10.1016/S0378-3758(02)00302-6

Bouza Herrera, C. N., & Al-Omari, A. I. (2011). Ranked set estimation with imputation of the missing observations: the median estimator. *Revista Investigación Operacional, 1* (32), 30-37.

Bouza, C. (2001). Model assisted ranked survey sampling. *Biometrical Journal,* (36), 753-764. https://doi.org/10.1002/1521-4036(200105)43:2<249::aid-bimj249>3.0.co;2-u

Chen, Z. B. (2004). Ranked set sampling: theory and applications. *Lectures Notes in Statistics, 176*, Springer, N. York. https://doi.org/10.1007/978-0-387-21664-5_7

David, H. A. (2003). Order Statistics, Third Edition, Hoboken. *New Jersey: John Wiley and sons, Inc.* https://doi.org/10.1002/0471722162

McIntyre, G. (1952). A method for unbiased selective sampling using ranked sets. *Australian Journal Agricultural Research*, (3), 385-390. https://doi.org/10.1071/AR9520385

Singh, H. P., & R. T. (2010). On ratio and product methods with certain known population parameters of auxiliary variable in sample surveys. *SORT*, (34), 157-180.

Takahasi, K. A. (1968). On the unbiased estimates of the population mean based on the sample stratified by means of ordering. *Annals of the Institute of Statistical Mathematics,* (20), 1-31. https://doi.org/10.1007/BF02911622

Copyrights