RANKED SET SAMPLING PROCEDURES FOR THE ESTIMATION OF THE POPULATION MEAN UNDER NON RESPONSES: A COMPARISON

Carlos N. Bouza[1]
Facultad de Matematica y Computacion
Universidad de La Habana
San Lázaro y L. Habana 10 400, Cuba.

ABSTRACT:
This paper analyses the use of ranked set sampling procedures for obtaining the sub-sample from the set of non-respondents. The first visit may serve for ranking accurately the sub-sampled non-respondents. The usual ranked set sample (rss) design and two variations of it : extreme-rss and median-rss are used for developing estimators of the population mean. Their expected variances and biases are obtained. A Monte Carlo experiment is developed for evaluating the behavior of the estimators. The use of rss appears as the best alternative.

KEY WORDS: extreme ranked set sampling, median ranked set sampling, non-response stratum, expected error.

MSC: 62D05

RESUMEN
En este trabajo se analiza el uso del muestro por rangos ordenados (ranked set sampling) para obtener la sub-muestra del conjunto de los no-respondientes.
La primera visita puede servir para rankear adecuadamente los sub-muestreados. El diseño rss usual tiene dos variaciones de este : extremal-rss y mediana-rss son usados para desarrollar estimadores de la media de la población . Sus varianzas esperadas y sesgos son obtenidos. Un experimento de Monte Carlo se desarrolla para evaluar el comportamiento de los estimadores. El uso de rss aparece como la mejor alternativa.

1 INTRODUCTION

The usual theory of survey sampling is developed assuming that the finite population $U=\{u_1 ,\ldots,u_N \}$ is composed by individuals that can be perfectly identified . A sample s of size $n \leq N$ is selected. The variable of interest Y is measured in each selected unit. Real life surveys should deal with problems that invalidate some initial assumptions and affect the properties of the statistical models. One of them appears when some of the units in the sample (responding units) do not give a response. The existence of non-responses do not permit to compute the sample mean

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \tag{1.1}$$

which estimates the population mean μ because we obtain response only from the units in

$.s_1=\{i \in s \mid i \text{ gives a response at the first visit}\}$

This fact suggests that the population U is divided into two strata: U_1 , where are grouped the units that give a response at the first visit, and U_2 which contains the rest of the individuals. This is the so called 'response strata' model and was first proposed by Hansen-Hurvitz (1946), see Cochran (1977). They proposed to select a subsample s_2' of size n'_2 among the n_2 non-respondents grouped in the sample s_2 . Then we obtain information on the non-respondent's strata U_2 through s_2' .

In this paper we consider the use of different rss schemes for selecting the subsample among the non-respondents. Rss was proposed by McIntyre (1952) using practical evidence. He claimed that rss produced more accurate estimators of the sample mean than the usual srswr design. Takahasi-Wakimoto

[1] bouza@matcom.uh.cu

(1968) gave mathematical support to his claims. Dell-Clutter (1972) established that even if the ranking is not perfect the proposed estimator is still unbiased. The use of rss is the theme a growing number of papers. Patil et.al. (2002) gave a review of the theme as well as a large list of papers.

Generally the first visit allows obtaining information on Y from each non-respondent. Hence we may use it for ranking the units in the subsample s'_2 and use a ranked set sampling (rss) method for selecting the sample of the units to be revisited. The performance of rss as a better alternative than simple random sample with replacement (srswr), in terms of accuracy, has been obtained under different variants of it, see Muttlak (1996) and (1997) and Samawi et.al. (1996) .

We develop estimators of the population mean under rss desings. Their errors are obtained. The behavior of them is compared using data provided by two large studies. One of them is related with the study of forest bio-diversity in mountains. All the selected units were visited but some of them were deficiently evaluated. Then a second visit to all those sites was made. The first visit provided sufficient information for ranking the biodiversity index of the non-respondents. The second research was related with the estimation of the percent of insects in sugar cane fields. Though they were visited an imperfect count of eggs, larvae and adults was detected. Then a subsample of them was revisited but an idea of the percentages was obtained from the first evaluation. Then the ranking in both does not involve an additional investment.

In Section 2 we present the basic results on the use of srswr for estimating the population mean under non-responses. Section 3 develops the rss counterparts using the classic selection procedure, extreme rss and median rss. Section 4 presents a study of the behavior of the different models. Monte Carlo experiments were designed for selecting samples, with each design and evaluating the accuracy of the estimator.

2 THE NON- RESPONSE STRATUM APPROACH UNDER SRSWR

Non responses may be motivated by a refusal of some units to give the true value of Y or by other causes. Refusals to respond are present in the majority of the surveys. Hansen-Hurvitz in 1946 proposed to select a sub-sample among the non-respondents, see Cochran (1977). This feature depends heavily on the proposed sub-sampling rule. Alternative sampling rules to Hansen-Hurvitz´s rule have been proposed see for example Srinath (1971) and Bouza (1981). It is described as follows:

Step 1: Select a sample s from U using srswr
Step 2: Evaluate Y among the respondents and determine $\{y_i : i\in s_1\subset U_1, |s_1| = n_1\}$.

Compute $$\bar{y}_1 = \frac{\sum_{i=1}^{n_1} y_i}{n_1} \qquad (2.1)$$

Step 3: Determine $n_2' = n_2/K$, $K>1$; $|s_2| = n_2$ with $s_2 = s - s_1$.
Step 4. Select a sub-sample s'_2 of size n_2' from s_2 using srswr.
Step 5. Evaluate Y among the units in s_2' $\{y_i : i\in s_2' \subset s_2, s_2\subset U_2\}$.

Compute $$\bar{y}'_2 = \frac{\sum_{i=1}^{n'_2} y_i}{n_2} \qquad (2.2)$$

Step 6. Compute the estimate of μ

$$\bar{\bar{y}} = \frac{n_1}{n}\bar{y}_1 + \frac{n_2}{n}\bar{y}'_2 = w_1\bar{y}_1 + w_2\bar{y}'_2 \qquad (2.3)$$

Note that (2.1) is the mean of a srswr-sample selected from $U_{1,}$ then its expected value is the mean of Y in the respondent stratum: μ_1. We have that the conditional expectation of (2.2) is:

$$E[\bar{y}'_2 | s] = \bar{y}_2 \qquad (2.4)$$

(2.4) is the mean of a srswr-sample selected from U_2 then

$$EE[\bar{y}'_2 \mid s] = \mu_2 \qquad (2.5)$$

Taking into account that for i=1,2 $E(n_i)=nN_i/N=nW_i$ the unbiasedness of (2.3) is easily derived, see Cocrhan (1977).

The variance of (2.3) is deduced by using the following trick;

$$\overline{\overline{y}} = (w_1\bar{y}_1 + w_2\bar{y}_2) + w_2(\ \bar{y}'_2 - \bar{y}_2) \qquad (2.6)$$

The first term is the mean of s, then its variance is σ^2/n. For the second term we have that

$$V\left(w_2(\bar{y}'_2 - \bar{y}_2)\middle| s\right) = w^2{}_2 E\left(\bar{y}'_2 - \mu_2) - (\bar{y}_2 - \mu_2)\middle| s\right)^2 =$$
$$= \quad w^2{}_2[E\left(\bar{y}'_2 - \mu_2)\middle| s\right)^2 + E\left((\bar{y}_2 - \mu_2)\middle| s\right)^2 \quad - 2E\left(\bar{y}'_2 - \mu_2)((\bar{y}_2 - \mu_2))\middle| s\right)]$$

Conditioning to a fixed n_2 we have that the expectation of the third term is
$(\bar{y}_2 - \mu_2)^2$. Then we have that:

$$V\left(w_2(\bar{y}'_2 - \bar{y}_2)\middle| s\right) = w^2{}_2\left(\frac{\sigma_2{}^2}{n'_2} - \frac{\sigma_2{}^2}{n_2}\right) = w^2{}_2\sigma_2{}^2\left(\frac{K}{n_2} - \frac{1}{n_2}\right) \qquad (2.7)$$

and

$$EV\left(w_2(\bar{y}'_2 - \bar{y}_2)\middle| s\right) = \frac{W_2(K-1)\sigma_2{}^2}{n} \qquad (2.8)$$

Hence the expected error of (2.3) is given by the well known expression

$$EV(\overline{\overline{y}}) = \frac{\sigma^2}{n} + \frac{W_2(K-1)\sigma_2^2}{n} \qquad (2.9)$$

3 THE USE OF THE RSS FOR SUBSAMPLING S_2.

Our proposal is to use a rss procedure for sub-sampling s_2. We take a subsample $s'_{2(rss)}$ from s using rss procedure. That is we select n'_2 independent samples of size $n'_2=n_2/K$ using srswr. The units are ranked accordingly with the variable close related with the variable of interest Y. Let
be the n'_2 independent random samples

$$Y_{11}, Y_{12} ...., Y_{1n'_2}; Y_{21}, Y_{22} ...., Y_{2n'_2}; ....; Y_{n'_2 1}, Y_{n'_2 2} ...., Y_{n'_2 n'_2}$$

They are ranked and we obtain

$$Y_{(1:1)}, Y_{(2:1)} ...., Y_{(n'_2:1)}; Y_{(1:2)}, Y_{(2:2)} ...., Y_{(n'_2:2)}; ....; Y_{(1:n'_2)}, Y_{(2:n'_2)} ...., Y_{(n'_2:n'_2)}$$

Where $Y_{(j:t)}$ is the j-th order statistics (os) of the t-th sample, j=1,…,n'_2 and t=1,…,n'_2.
The rss sample is formed by the n'_2 os in the diagonal. That is the measurements of Y are

$$Y_{(1:1)}, Y_{(2:2)} ...., Y_{(n'_2:n'_2)}$$

The estimate of μ_2 is made by using the estimator:

$$\bar{y}'_{2(rss)} = \frac{\sum_{j=1}^{n'_2} Y_{(j:j)}}{n'_2} \qquad (3.1)$$

Note that $E[Y_{(j:j)} \mid n_2] = \mu_{(j)}$, $j=1,\ldots,n'_2$. At this randomization stage the parameter is the mean of y in s_2.

Hence

$$E\left(\bar{y}'_{2(rss)}\right) = E\left(\frac{\sum_{j=1}^{n'_2} E\left(Y_{(j:j)}\right)}{n'_2}\right) = E\left[\bar{y}_2\right] = \mu_2$$

The rss counterpart of (2.3) is

$$\bar{y}_{(rss)} = \frac{n_1}{n}\bar{y}_1 + \frac{n_2}{n}\bar{y}'_{2(rss)} = w_1\bar{y}_1 + w_2\bar{y}'_{2(rss)} \qquad (3.2)$$

It can be represented by

$$\bar{\bar{y}}_{(rss)} = \left(w_1\bar{y}_1 + w_2\bar{y}_2\right) + w_2\left(\bar{y}'_{2(rss)} - \bar{y}_2\right)$$

Its conditional variance is

$$V(\bar{\bar{y}}_{(rss)} | s) = \frac{\sigma^2}{n} + w_2^2 V\left(\bar{y}'_{2(rss)} - \bar{y}_2 | s\right)$$

We need to obtain an explicit expression of the second term in the right hand side. It is:

$$V\left(w_2\left(\bar{y}'_2 - \bar{y}_2\right) | s\right) = w^2{}_2 E\left(\bar{y}'_{2(rss)} - \mu_2) - (\bar{y}_2 - \mu_2) | s\right)^2 =$$

$$= \quad w^2{}_2[E\left(\bar{y}'_{2(rss)} - \mu_2) | s\right)^2 + E\left((\bar{y}_2 - \mu_2) | s\right)^2 \quad - 2E\left(\bar{y}'_{2(rss)} - \mu_2)\left((\bar{y}_2 - \mu_2)\right) | s\right)]$$

The first term of the equation within brackets is equal to

$$E\left(\bar{y}'_{2(rss)} - \mu_2) | s\right)^2 = \frac{\sum_{j=1}^{n'_2} \sigma^2_{(j)}}{n'_2} = \frac{\sigma_2^2}{n'_2} - \frac{\sum_{j=1}^{n'_2} \Delta^2_{(j)}}{n'_2}$$

Where

$$\frac{\sum_{j=1}^{n'_2} \Delta^2_{(j)}}{n'_2} = \frac{\sum_{j=1}^{n'_2} (\mu_{(j)} - \mu)^2}{n'_2}$$

The second term is related to the use of srswr for selecting s_2 and it is equal to

$$E\left((\bar{y}_2 - \mu_2) | s\right)^2 = \frac{\sigma_2^2}{n_2}$$

and

$$E\left(E[(\bar{y}'_{2(rss)}-\mu_2)]((\bar{y}_2-\mu_2))\middle|s\right)=E\left(((\bar{y}_2-\mu_2))^2\middle|s\right)=\frac{\sigma_2^2}{n_2}$$

Hence the counterpart of (2.7) is

$$V\left(w_2\left(\bar{y}'_{2(rss)}-\bar{y}_2\right)\middle|s\right)=w^2{}_2\left(\frac{\sigma^2_{2(rss)}}{n'_2}-\frac{\sigma_2{}^2}{n_2}\right)=w^2{}_2\left(\frac{\sigma_2^2}{n'_2}-\frac{\sigma_2{}^2}{n_2}-\frac{\sum_{j=1}^{n'_2}\Delta^2_{(j)}}{n'_2}\right) \qquad (3.4)$$

Substituting $n'_2=n_2/K$ we have that the two first terms are equal to (2.7). Hence we have that

$$EV(\overline{\overline{y}}_{(rss)})=\frac{\sigma^2}{n}+\frac{W_2(K-1)\sigma_2^2}{n}-W_2E\left(\frac{K\sum_{j=1}^{n'_2}\Delta^2_{2(j)}}{n}\right) \qquad (3.5)$$

which is smaller than (2.9) because the last term is positive.

Some variations of the basic rss procedure have been proposed. They have a practical sound basis because in some occasions to rank all the units may be subject to large errors. Detecting only some units with distinguished ranks may be easier and accurate. Take for example the identification all the two extremes values $Y_{(1:j)}$ and $Y_{(n:j)}$ in the j-th sample. Samawi et.al. (1996) proposed this rss procedure named extreme rss (erss). This procedure considers the identification of the extremes in the samples .

Considering that n'_2 is even we evaluate only some extremes

$$Y_{2(j:e)}=\begin{cases}Y_{2(j:1)} & if \quad j=1,\ldots,\frac{n'_2}{2}\\ Y_{2(j:n'_2)} & if \quad j=\frac{n'_2}{2}+1,\ldots,n'_2\end{cases}$$

An estimator of μ_2 is:

$$\bar{y}'_{2(erss)}=\frac{\sum_{j=1}^{n'_2}Y_{2(j:e)}}{n'_2}=\frac{Y_{2(1)}+Y_{2(n'_2)}}{2} \qquad (3.6)$$

We have that

$$E(Y_{2(j:e)})=\begin{cases}\mu_{2(1)} & if \quad j=1,\ldots,\frac{n'_2}{2}\\ \mu_{2(n'_2)} & if \quad j=\frac{n'_2}{2}+1,\ldots,n'_2\end{cases}$$

The estimator is biased because

$$E(\bar{y}'_{2(erss)}) = \frac{\mu_{2(1)} + \mu_{2(n'_2)}}{2} = \mu_{2(e)}$$

Note that it is different from μ_2 but, if the distribution is symmetric with respect to μ, its bias

$B(erss)=\mu_{2(e)} - \mu_2=[(\mu_{2(1)} - \mu_2)+(\mu_{2(n'2)} - \mu_2)]/2$ (3.7)

is equal to zero. Then the symmetry of the distribution plays a role in the magnitude of the bias. The variance of the involved os´s is:

$$V(Y_{2(j:e)}) = \begin{cases} V(Y_{2(j:1)}) = \sigma^2_{2(1)} & if \quad j = 1,\ldots,\frac{n'_2}{2} \\ V(Y_{2(j:n'_2)}) = \sigma^2_{2(n'_2)} & if \quad j = \frac{n'_2}{2}+1,\ldots,n'_2 \end{cases}$$

Then

$$V(\bar{y}'_{2(erss)}) = \frac{\sum_{j=1}^{n'_2} V(Y_{2(j:e)})}{n'^2_2} = \frac{\sigma^2_{2(1)} + \sigma^2_{2(n'_2)}}{2n'_2} = \frac{\sigma^2_2}{n'_2} - \frac{\Delta^2_{2(1)} + \Delta^2_{2(n'_2)}}{2n'_2}$$

Is the conditional variance and we have that

Hence we have established hat for estimating the mean in the population we may use the estimator

$$\bar{y}_{(erss)} = w_1\bar{y}'_{2(erss)} = w_1\bar{y}_1 + w_2\bar{y}'_{2(erss)} \qquad (3.8)$$

Using (3.7) we have that its bias is $W_2B(erss)$

Taker $\Delta_{2(1)} = \mu_{2(1)} - \mu_2$ and $\Delta_{2(n'_2)} = \mu_{2(n'_2)} - \mu_2$ its expected variance is given by :

$$EV(\bar{\bar{y}}_{2(erss)}) = \frac{\sigma^2}{n} + \frac{W_2(K-1)\sigma^2_2}{n} - \frac{W_2K\left(\Delta^2_{2(1)} + E(\Delta^2_{2(n'_2)})\right)}{2n}$$

The third term, at the right hand side of the equation, is the gain in accuracy due to the use of the rss method proposed by Samawi et.al. (1996) with respect to the srswr model.
A preference for rss when compared with erss is obtained when :

$$E\left(\sum_{j=1}^{n'_2} \Delta^2_{2(j)}\right) > E\left(\frac{\Delta^2_{2(1)} + E(\Delta^2_{2(n'_2)})}{2}\right)$$

Another modification of rss is to use only the median of each rss sample. That is, we compute only the os corresponding to the median of the j-th sub-sample. As we assumed that n'_2 is even we take as median

$$Y_{2(j:med)} = \begin{cases} Y_{2(j:\frac{n'_2}{2})} & if \quad j = 1,2,\ldots,n'_2/2 \\ Y_{2(j:\frac{n'_2+2}{2})} & if \quad j = \frac{n'_2}{2}+1,\ldots,n'_2 \end{cases}$$

The estimator of the mean of μ_2 when the median rss (mrss) procedure is used is:

$$\bar{y}'_{2(mrss)} = \frac{\sum_{j=1}^{n'_2} Y_{2(j:med)}}{n'_2} \qquad (3.9)$$

This is also a biased estimator because

$$E(\bar{y}'_{2(mrss)}) = \frac{\mu_{2(\frac{n'_2}{2})} + \mu_{2(\frac{n'_2+2}{2})}}{2} = \mu_{2(m)}$$

When we deal with distributions symmetric with respect to μ we may expect that if will be close to $\mu_{2(m)}$. A good example is the normal distribution where the median and mean coincides. In general the bias of (3.9) is:

$$B(mrss)) = \frac{\left(\mu_{2(\frac{n'_2}{2})} - \mu\right) + \left(\mu_{2(\frac{n'_2+2}{2})} - \mu\right)}{2}$$

The variance of the involved random variable is given by:

$$V(Y_{2(j:m)}) = \begin{cases} V(Y_{2(j:\frac{n}{2})}) = \sigma^2_{2(\frac{n'_2}{2})} & if \quad j = 1,\ldots,\frac{n'_2}{2} \\ V(Y_{2(j:\frac{n'_2+2}{2})}) = \sigma^2_{2(\frac{n'_2+2}{2})} & if \quad j = \frac{n'_2}{2}+1,\ldots,n'_2 \end{cases}$$

Hence

$$V(\bar{y}'_{2(mrss)}) = \frac{\sum_{j=1}^{n'_2} V(Y_{2(j:m)})}{n'^2_2} = \frac{\sigma^2_{2(\frac{n'_2}{2})} + \sigma^2_{2(\frac{n'_2+2}{2})}}{2n'_2} = \frac{\sigma^2_2}{n'_2} - \frac{\Delta^2_{2(\frac{n'_2}{2})} + \Delta^2_{2(\frac{n'_2+2}{2})}}{2n'_2} = \frac{\sigma^2_2}{n'_2} - \frac{\Delta^2_{2(m)}}{2n'_2}$$

Note that if B(mrss) is negligable then $\Delta_{2(m)}$,which measures the gain in accuracy with respect to srs, is also small.

Mimicking the construction of the other estimators of μ we have for mrss

$$\bar{\bar{y}}_{(mrss)} = w_1\bar{y}_1 + w_2\bar{y}'_{2(mrss)} \qquad (3.10)$$

This estimator's bias B(mrss) is negligable if the data are approximately symmetric with respect to the median and its expected variance is:

$$EV(\bar{\bar{y}}_{2(mrss)}) = \frac{\sigma^2}{n} + \frac{W_2(K-1)\sigma^2_2}{n} - \frac{W_2KE\left(\Delta^2_{(m)}\right)}{2n}$$

Rss should be preferred if

$$E\left(\sum_{j=1}^{n'_2} \Delta^2_{2(j)}\right) > E\left(\frac{\Delta^2_{2(m)}}{2}\right)$$

$$E\left(\Delta^2_{(m)}\right) > E\left(\frac{\Delta^2_{2(1)} + E(\Delta^2_{2(n'_2)})}{2}\right)$$

Erss is better then mrss when

4. A MONTE CARLO COMPARISON OF THE ACCURACY OF THE MODELS.

We used two data base sets . They were considered as providing the set of values of the interest variable Y in the population: $Y_1,\ldots Y_N$. Some of the Y_j's are identified as non-respondents. They correspond to units for which the first measurement was inaccurate and a second visit was made for obtaining a correct evaluation.. Hence, once a sample s was selected we were able to identify s_1 and s_2. In our notation each rss procedure is identified with :

R=rss, erss, mrss.

The Monte Carlo experiment worked as follows:

Step 1. We select s then the sample mean of Y in s_1 is calculated and n'_2 is determined.
Step 2. We select n'_2 sub-samples from s_2 and they are ranked.
Step 3. A Bootstrap procedure selects re-samples of size n'_2 using srswr from each of the n'_2 sub-samples.
Step 4. For each b+1,..,B the Bootstrap estimate of μ :

$$\overset{=}{y}_{(R)mb}\ w_1\bar{y}_1 + w_2\bar{y}'_{(2R)b}$$

.is computed for the m-th sample using (3.2) and (3.8) correspondingly to R.

The cycle is repeated for obtaining M samples. Then the variance is estimated and the Bootstrap confidence interval (CI) is calculated using the B obtained Bootstrap's samples. As we know the real value of μ we can compute the proportion of times that the CI contains it. R identifies the rss estimator to be used for estimating the non-respondent's stratum means.

The Bootstrap procedure algorithm used is described as follows.

Bootstrap Procedure

Fix $Y=\{Y_1,\ldots, Y_N\}$, K, M and B.
While m<M do
.m=0, h=0, π(R)=0
Select a sample $\{y_1,\ldots,y_n\}$from Y using simple random sampling with replacement.
If y_j is a non-respondent then $y_j \in s_2$, $|s_2|=n_2$, $|\{j \notin s_2\}|=n_1$, $n_2'=\lfloor n_2/K \rfloor$
.$w_1=n_1/n$, $w_2=n_2/n$

Compute

$$\bar{y}_1 = \frac{\sum_{j\notin s_2} y_j}{n_1}$$

While b<B do
While h<n'_2 do
Select a sample $s_{2h}=\{y_1,\ldots,y_{n2'}\}$from s_2 using simple random sampling with replacement.
Rank s_{2h} and determine the ranked sample $s_2(h)$
.h=h+1
Select using srswr a Bootstrap subsample s_{2hb} from s_{2h}
Compute

$$\bar{y}_{(R)b} = \frac{\sum_{j=1}^{n'_2} y_{(j:R)b}}{n'_2}$$

$$\bar{\bar{y}}_{(R)mb}\ w_1\bar{y}_1 + w_2\bar{y}'_{(2R)b}$$

.b =b+1

. Calculate

$$\bar{\bar{y}}_{(R)mB} = \frac{\sum_{b=1}^{B} \bar{\bar{y}}_{(R)mb}}{B}$$

$$s_{(R)mB} = \sqrt{\frac{\sum_{b=1}^{B}\left(\bar{\bar{y}}_{(R)mb} - \bar{\bar{y}}_{(R)mB}\right)^2}{B-1}}$$

$$I_{(R)mB} = \left(\bar{\bar{y}}_{(R)mB} - \frac{2s_{(R)mb}}{\sqrt{B}}, \bar{\lambda} + \frac{2s_{(R)mB}}{\sqrt{B}}\right)$$

$$Z_{(R)m} = \begin{cases} 1 & if \quad \mu \in I_{(R)mB} \\ 0 & otherwise \end{cases}$$

$$\pi(R) = \pi(R) + Z_{(R)m}$$

M=M+1

$\pi(R)_{)}$ =π(R)/M

END

Note that the CI uses 2 as an approximation of the 95% percentile.

We used K=2, 5 and 10, B/n≅ 0.1, 0.2 and 0.5 f=n/N≅0.1, 0.05 and 0.01 and M=100. Considering the proportions ρ(R), the relative evaluation of a method's precision is measured by:

$\rho(R)=\sum^{M}_{m=1}|\mu_{(R)} - \mu|_{m}/M\mu$

where μ(R) is the estimator of the mean μ made by the corresponding rss estimator.

Table 4.1 : Percent of Coverage of the Confidence Intervals: 100π(R) for the variable Y=Coefficient of Infestation in Sugar Cane fields

B/n≅0.1									
Subsample parameter	.rss			erss			.mrss		
	.f=0.1	f=0.05	f=0.01	.f=0.1	f=0.05	f=0.01	.f=0.1	f=0.05	f=0.01
K=2	96.8	93.2	92.5	89.4	83.7	84.3	81.5	79.5	77.6
.K=5	94.2	89.5	91.0	84.4	81.7	81.8	81.2	78.5	73.5
K=10	94.1	89.6	91.0	84.7	80.8	81.3	81.4	77.7	71.9
B/n≅0.2									
K=2	96.9	93.3	92.8	89.3	83.4	85.2	82.2	80.1	77.8
.K=5	94.4	89.3	91.2	84.0	81.4	81.4	81.7	78.8	73.2
K=10	94.3	89.2	90.4	84.1	80.3	81.7	81.5	77.9	71.7
B/n≅0.5									
K=2	96.8	93.3	92.1	89.1	83.4	84.4	81.7	79.9	77.5
.K=5	94.2	89.3	91.1	84.0	81.8	81.9	81.2	77.5	73.7
K=10	94.1	89.2	91.0	84.1	80.9	81.4	81.9	77.3	71.2

Table 4.2 : Percent of Coverage of the Confidence Intervals: 100π(R) for the variable Y= Hemoglobin in blood in Adolescents

Subsample parameter	.rss			erss			.mrss		
	.f=0.1	f=0.05	f=0.01	.f=0.1	f=0.05	f=0.01	.f=0.1	f=0.05	f=0.01
B/n≅0.1									
K=2	95.5	93.2	90.0	89.4	83.7	84.3	94.3	92.0	89.4
.K=5	94.7	91.4	93.6	84.4	81.7	81.8	93.6	91.0	87.4
K=10	94.3	91.1	90.4	84.7	80.8	81.3	92.9	94.5	85.0
B/n≅0.2									
K=2	93.4	92.9	92.9	89.3	83.4	85.2	97.5	95.1	92.0
.K=5	92.4	90.0	90.6	84.0	81.4	81.4	91.1	93.7	91.7
K=10	92.2	94.3	93.0	84.1	80.3	81.7	92.3	91.4	88.4
B/n≅0.5									
K=2	95.4	90.0	90.6	89.1	83.4	84.4	94.8	92.7	92.9
.K=5	95.3	94.4	93.1	84.0	81.8	81.9	94.3	89.6	91.6
K=10	94.7	94.0	90.3	84.1	80.9	81.4	95.8	90.5	91.7

Table 4.2 presents the percentage of coverage of μ by the Bootstrap CI's computed using samples from the data providing from Hemoglobin's analysis. Again the use of rss is the best option but mrss has a good behavior for f=0.1 and f=0.05 as well as when B/n≅0.5. The increase in this parameter is generally associated with better values of π(mrss). These results may be generated by the fact that the percent in hemoglobin is well described by a normal distribution. The behavior of erss again is poor.

Table 4.3 : Values of ρ(R) for the variable Y=Coefficient of Infestation in Sugar Cane fields

Subsample parameter	.rss			erss			.mrss		
	.f=0.1	f=0.05	f=0.01	.f=0.1	f=0.05	f=0.01	.f=0.1	f=0.05	f=0.01
K=2	0.43	0.48	0.52	0.99	0.96	0.94	0.42	0.42	0.44
.K=5	0.51	0.50	0.56	0.99	0.92	0.87	0.43	0.57	0.59
K=10	0.56	0.53	0.56	0.97	0.92	0.98	0.49	0.56	0.59

A look to table 4.3 suggests that for rss the increment of f and a diminishing in K have a significant influence in obtaining small values of ρ(rss). It seems that the levels of f and K have not a significant influence in ρ(rss). A similar comment may be made on the behavior of Erss. This procedure is considerably more inaccurate than rss. ρ(mrss) is always smaller than ρ(erss) for f=0.1 it performs better than rss for K=2.

Table 4.4 : Values of ρ(R) for the variable Y= Hemoglobin in blood in Adolescents

Subsample parameter	.rss			erss			.mrss		
	.f=0.1	f=0.05	f=0.01	.f=0.1	f=0.05	f=0.01	.f=0.1	f=0.05	f=0.01
K=2	0.27	0.22	0.29	0.67	0.62	0.63	0.21	0.21	0.26
.K=5	0.37	0.29	0.34	0.80	0.72	0.73	0.21	0.24	0.21
K=10	0.42	0.31	0.31	0.94	0.84	0.85	0.20	0.22	0.26

The results given in table 4.4 suggest that for rss the increment of K determines larger value of ρ(rss). It seems that the levels of f have not a significant influence in ρ(rss). Erss has a worse behavior compared with the other procedures. Its accuracy is seriously affected by the increments in K and f. Mrss has a better behavior than rss which is not seriously affected by changes in any of the parameters. Again the possible normality of the involved variable should be having a determinant influence in the behavior of the accuracy of mrss.

Acknowledgements: The author acknowledges the support given by the Third World Academy of Sciences to the project where this research was developed.

RECEIVED AUGUST, 2006
REVISED FEBRUARY, 2007

REFERENCES

[1] BOUZA, C.N. (1981): Sobre el problema de la fracciòn de muestreo para el caso de las no respuestas. **Trabajos de Estadística**. 21, 18-24

[2] COCHRAN, W.G. (1977): **Sampling techniques**. Wiley, N. York-

[3] DELL, T.R. and CLUTTER, J.L. (1972): Ranked set sampling theory with order statistics. Background. **Biometric**s 28, 545-555.

[4] MCINTYRE, G.A. (1952): A method of unbiased selective sampling using ranked sets. **Australian J. Agricultural Research**. 3, 385-390.

[5] MUTTLAK, H.A. (1996) Median Ranked set sampling . **J. of Applied Statistical Science**. 6, 91-98.

[6] PATIL, G.P (2002): Ranked set sampling. In **Encyclopedia of Enviromentrics**. (A.H. El-Shaarawi and W.W. Pieegoshed, (Editors). Vol. 3,1684-1690. Wiley, Chichester.

[7] SAMAWI, H. ABU-DAYYEH, W. AND AHMED, S. (1996): Extreme ranked set sampling. **Biometrical J.** 30, 577-586-

[8] SRINATH, K. P. (1971): Multi-phase sampling in non-response problems**. J. American Stat. Ass**. 66, 583-586.

[9] TAKAHASi K. and WAKIMOTO, K. (1968): On unbiased estimates of the population mean based on sample stratified by means of ordering. **Annals of the Inst. of Statistical Mathematics**. 20, 1-31