OPTIMAL ASPECTS OF A MODEL BASED RANDOMIZED RESPONSES PROCEDURE UNDER UNEQUAL SELECTION OF INSENSITIVE VARIABLES

Carlos N. Bouza*[1] and Lakshma Singh**
* Facultad de Matemática y Computación, Universidad de La Habana, Cuba
**Bhat & Sarkar Informatic Consultors, India

RESUMEN:
En este trabajo se presentan algunos resultados recientemente desarrollados en el estudio de los modelos de respuestas aleatorizadas. El uso del muestreo de conjuntos ordenados es analizado.

ABSTRACT:
En este trabajo se presentan algunos resultados recientemente desarrollados en el estudio de los modelos de respuestas aleatorizadas. El uso del muestreo de conjuntos ordenados es analizado.

KEY WORDS: ranked set sampling, random response query, gain in accuracy

MSC: 62G05

1. INTRODUCTION

Assume that Y is a sensitive variable which we plan to evaluate in a finite population $U=\{u_1,\ldots,u_N\}$. Some individuals have values of Y that carries a stigma. Hence he/she will tend to give incorrect information or to refuse to answer. It is well known that when we deal with sensitive questions we should face the need to reduce the refusals to respond and the response bias. A possibility is to replace the direct response to the sensitive question by using a random response (RR) query. The seminal work is due to Warner (1965). Some recent contributions on the theme are Singh-Singh (1993) and Zou (1997) for example. All the papers on RR have a common feature: a traditional sampling design provides the sample.

Warner's RR-model dealt with a qualitative question: the possible responses are ´yes´ or ´not´. The aim of the surveyor is to estimate the probability of having the stigma. It is expected that a large percent of the persons bearing the stigma will lie or refuse to answer. The basic Warner's method consists in placing the question associated with the stigma together with some insensitive ones. The respondent chooses randomly a question and answers it without revealing which was selected. When we deal with a quantitative character a similar reasoning can be used. Chaudhuri-Stenger (2005) developed a model for obtaining a report which is the evaluation of a function of the sensitive variable and of other innocuous ones.

.After the seminal paper of Warner (1965) different contributions have generated a large set of models. The models are generally based on the selection a sample using simple random sampling with replacement.. The collection of models is being enlarged with recent contributions as Singh-Singh (1993), Chaudhuri et. al. (1996), Zou (1997) and Singh et. al. (1998). Singh et.al. (1998) proposed the use of a randomized

[1] bouza@matcom.uh.cu

procedure for estimating the mean of a quantitative character and the proportion of members of a gang in the population. Two independent samples should be selected by using srswr.

Ranked set sampling (rss) was first proposed by McIntire (1952). He used this model for estimating the mean of pasture yields. This design appeared as a useful technique for improving the accuracy of the estimation of means. This fact was affirmed but a mathematical proof of it was settled by Takahashi-Wakimoto (1968). In many situations the statistician deals with the need of combining some control and the implementation of some flexibility with the use of a random based sample. This is a common problem in the study of environmental and medical studies. In these cases the researcher generally has abundant and accurate information on the population units. It is related with the variable of interest Y and to rank the units using this information is cheap. The rss procedure is based on the selection of m independent samples, not necessarily of the same size, by using simple random sampling (srs). The sampled units are ranked and the selection of the units evaluated takes into account the order of them in the combined m samples. The proposal of McIntire (1952) was to use a prediction of Y. After some experiences with its application the lack of a coherent statistical theory appeared as an interesting theme of study to theoretical statisticians. The units can be ranked by means of a cheap procedure and then an order statistics is selected from each of the independent samples selected using srs with replacement (srswr). It turned out that the use of ranked set sampling is highly beneficial and leads to an estimator which is more precise than the usual sample mean per unit. The method is now referred to as ranked set sampling (rss) method in the literature. See Patil (2002) and Patil et. al. (1994, 1999) for a detailed discussion . Some recent papers on the use of rss in sampling a finite population are Barabasi-Pisani (2002) and Bouza (2002 a, 2002b).

In this paper we will develop a study of the use of alternative rr procedures when rss is used instead of srs with replacement (srswr) for the above presented RR methods . Section 2 deals with the model for quantitative variables and section 3 with the qualitative variable case. We derive criteria for optimizing (maximizing) the gain in accuracy due to the use of rss.

2. CHAUDHURI-STENGER'S UNEQUAL PROBABILITY RR MODEL

2.1 The basic model

Let us describe briefly the RR procedure proposed by Chaudhuri-Stenger (2005). The procedure is implemented by determining a set of values $X=\{0,1,\ldots,T\}$. The interviewed person u_i performs a random experiment and selects a member of X which we denote by X_i. Take the probability of observing a certain value *as* $P(X_i=\lambda)=\pi_\lambda$, $\lambda=0,1,\ldots,T$. Assuming that it is a probability distribution :

$\sum^{T}_{\lambda=0} \pi_\lambda=1$

The report of an individual u_i is

$$W_i=\begin{cases} Y_i & if \quad X_i=0 \\ A_h & if \quad X_i=h \end{cases}$$

Hence $E_R(W_i)= \pi_0 Y_i+\sum^{T}_{h=1} \pi_h A_h$

Note that the mean of the insensitive variable A is known and is given by $\mu_A=\sum^{T}_{h=1} \pi_h A_h/\sum^{T}_{h=1} \pi_h$. Then we expect that the response of u_i be $\mu_{Wi}=\pi_0 Y_i+(1-\pi_0)\mu_A$. We can compute from the response of each u_i

$$S_i=\frac{W_i-(1-\pi_0)\mu_A}{\pi_0}$$

Under the described model M we have that $E_M(S_i)=Y_i$. Then S_i is model unbiased and its model variance is:

$$V_M(S_i) = \frac{\pi_0(1-\pi_0)Y_i^2 - 2\pi_0(1-\pi_0)Y_i\ \mu_A + \sum_{h=1}^{T}\pi_h A_h{}^2 - (1-\pi_0)\mu_A^2}{\pi_0{}^2} = V_i$$

When we select a sample s the procedure generates the data $D(S)=\{(u_i, W_i)/u_i \in s, W_i \in \{Y_i, A_1, \ldots, A_T\}\}$. The sample mean is:

$$\bar{S} = \frac{\sum_{I=1}^{n} S_i}{n} \qquad 2.2)$$

The model unbiasedness of S_i permits to derive easily that $E_d E_M(S_i)=\mu_Y$, where E_d is the design expectation. Hence we have that

$$E_d E_M(\bar{S}) = \mu_Y$$

We consider that the trials performed by each sampled u_i for given the report is independent of those made any other individual. Then the model variance of (2.2) is

$$V_M(\bar{S}) = \frac{\sum_{i=1}^{n} V_i}{n^2}$$

The expected model variance of (2.2) is

$$E_d(V_i) = \frac{\pi_0(1-\pi_0)\left(\sigma_Y^2 + \mu_Y^2\right) - 2\pi_0(1-\pi_0)\mu_A\mu_Y}{n^2\pi_0^2} + \Psi_1$$

where

$$\Psi_1 = \frac{\sum_{h=1}^{T}\pi_h A_h^2 - (1-\pi_0)^2\mu^2{}_A}{n\pi_0^2}$$

$$E_d V_M(\bar{S}) = \sum_{i=1}^{n} E_d\left[\frac{\pi_0(1-\pi_0)Y_i^2 - 2\pi_0(1-\pi_0)Y_i\ \mu_A + \sum_{h=1}^{T}\pi_h A_h{}^2 - (1-\pi_0)\mu_A^2}{n^2\pi_0{}^2}\right] \qquad (2.3)$$

The only random variable present in V_i are the functions of the sensitive variable Y_i. As $E_d(Y_i^2)=\sigma^2_Y + \mu^2_Y$

Theorem 2.1 . Consider the model based RR procedure described by the use of W_i given above. The expected value of the estimator (2.2) of the mean is unbiased and its expected sampling error is:

$$E_d V_M(\bar{S}) = V_{srs} = \frac{(1-\pi_0)\left(\sigma_Y^2 + \mu_Y^2\right) - 2(1-\pi_0)\mu_A\mu_Y}{n\pi_0} + \frac{\Psi_1}{n\pi_0^2}$$

When srswr is the sampling design used for selecting the sample

2.2. The rss alternative to Chaudhuri-Stenger's unequal probability RR model.

The basic rss procedure can be described as follows:

Procedure RSS1

While $t<m$ do

Select a sample unit independently from U using srswr.

Each unit in $s_{(t)}$ is ranked and the order statistics (os) $Y_{(1:t)},....,Y_{(r(t):t)}$ are determined.

END

Then the procedure generates the matrix

$\boldsymbol{Y_{(1:1)}}$	$Y_{(2:1)}$	• • •	$Y_{(t:1)}$	• • •	$Y_{(m:1)}$
$Y_{(1:2)}$	$\boldsymbol{Y_{(2:2)}}$	• • •	$Y_{(t:2)}$	• • •	$Y_{(m:2)}$
•	•	• • •	•	• • •	•
•	•	• • •	•	• • •	•
•	•	• • •	•	• • •	•
$Y_{(1:t)}$	$Y_{(2:t)}$	• • •	$\boldsymbol{Y_{(t:t)}}$	• • •	$Y_{(m:t)}$
•	•	• • •	•	• • •	•
•	•	• • •	•	• • •	•
•	•	• • •	•	• • •	•
$Y_{(1:m)}$	$Y_{(2:m)}$	• • •	$Y_{(t:m)}$	• • •	$\boldsymbol{Y_{(m:m)}}$

The ranked set sample is composed by the elements in the diagonal $s(i)=\{Y_{(i:i)}, i=1,..,m\}$. The procedure is repeated r times and the r samples determines a sample of size $n=rm$. The usual estimator of μ_Y en srswr is $\mu_{srs}=\sum_{i=1}^{n} Y_i/n$ and its variance is given by $V[\sum_{i=1}^{n} Y_i/n]=\sigma^2/n$. If we base our inferences on the os's the mean of them is:

$$\mu_{rss}=\sum_{t=1}^{r}\sum_{i=1}^{m} Y_{(i)t}/rm \qquad (2.4)$$

where $Y_{(i)t}$ is the ith-os of the ranked sample $s(t)$. It is well known that $E(Y_{(i)t})=\mu_{(i)}, t=1,..,r$. As $\mu=\sum_{i=1}^{m}\mu_{(i)}/m$ it is unbiased and the variance is $V[\sum_{t=1}^{r}\sum_{i=1}^{m} Y_{(i)t}/rm]=\sum_{i=1}^{m}\sigma^2_{(i)}/rm^2$, Arnold et al (1992):

Note that the ranks do not intervene in the selection of the sample. We can define a map $g(u_i)$ such that it assigns to each sampled unit u_i a rank and only one. Each sampled unit may be ranked using g without measuring Y using some judgments. Say that the rank represents certain judgment on the value of Y. The first arising question is whether this ranking affects the behavior of a statistical procedure based in it. The first results in this theme considered that the rank was perfect, see McIntyre (1952), Takahasi-Wakimoto (1968). Dell-Clutter (1972) studied this problem and derived that the unbiasedness of the estimator is maintained though an auxiliary variable is used for the ranking and $\sum_{i=1}^{m}(\mu_{(i)}-\mu)=\sum_{i=1}^{m}\Delta_{(i)}=0$. These differences play an important role in rss because the variance of an os is given *by* $\sigma^2_{(i)}=\sigma^2-\Delta^2_{(i)}$. An extreme case is that in which none of the ranks assigned by judgment coincide with the true ones. Then the orders are considered as assigned by a random mechanism and $\Delta_{(i)}=0$ for any $i=1,..,n$. In this case rss design is equivalent to the srs design.

We will analyze the use of rss when the RR is used for obtaining the reports. The sampler ranks using his/her believe on the value of Y. The individual ranked in the i-th place of the ordered sample $s_{(i)}$ gives his/her report. The set of reports $\{W_{(1)t},\ldots,W_{(m)t}\}$ is obtained in each cycle $t=1,..,r$. Under the model we have that

$$E_M(S_{(i)t})=Y_{(i)}$$

hence the structure of (2.4) suggests to use the estimator

$$\bar{S}_{(rss)} = \frac{\sum_{t=1}^{r}\sum_{i=1}^{m} S_{(i)t}}{rm} \tag{2.5}$$

It is unbiased because

$$E_d E_M(\bar{S}_{(rss)}) = \frac{\sum_{i=1}^{m} \mu_{Y(i)}}{m} = \mu_Y$$

The model variance of the estimator is obtained by dealing with the fact that for each observation the variance of the i-th os of the sample *s(t)* is:

$$V_M(S_{(i)t}) = V_{(i)} = \frac{\pi_0(1-\pi_0)Y_{(i)t}^2 - 2\pi_0(1-\pi_0)Y_{(i)t}\mu_A + \sum_{h=1}^{T}\pi_h A_h{}^2 - (1-\pi_0)\mu_A^2}{\pi_0{}^2}$$

Then the model variance of (2.5) is

$$V_M\left(\bar{S}_{(rss)}\right) = \frac{1}{rm^2}\sum_{i=1}^{m} V_{(i)}$$

We should calculate de expectation of the model variance for rss design. The expectation of the first term in the numerator is

$$E_d\left(\pi_0(1-\pi_0)Y_{(i)t}^2\right) = \pi_0(1-\pi_0)\left(\mu_{Y_{(i)}}^2 + \sigma_{Y_{*ii)}}^2\right)$$

As

$$\sigma_{Y_{(i)}}^2 = \sigma_Y^2 - \left(\mu_{Y_{(i)}} - \mu_Y\right)^2$$

Then we have the following theorem

Theorem 2.2 . Consider the model based RR procedure described by the use of W_i given above. The expected value of the estimator (2.5) of the mean is unbiased and its expected sampling error is:

$$E_d V_M(\bar{S}_{(rss)}) = V_{(rss)} = \frac{(1-\pi_0)\left(\sigma_Y^2 - \dfrac{\sum_{i=1}^{m}\Delta_{Y(i)}^2}{m}\right)}{n\pi_0} + \frac{(1-\pi_0)\sum_{i=1}^{m}\mu_{Y_{(i)}}^2}{nm\pi_0} + \psi_2 \tag{2.6}$$

where

$$\psi_2 = \frac{m\left[\left(\sum_{h=1}^{T}\pi_h A_h^2\right) - (1-\pi_0)^2\mu_A^2 - 2\pi_0(1-\pi_0)\mu_A\mu_Y\right]}{rm^2\pi_0{}^2} = \frac{\sum_{h=1}^{T}\pi_h A_h^2}{n\pi_0{}^2} - \frac{(1-\pi_0)^2\mu_A{}^2}{n\pi^2{}_0} - \frac{2(1-\pi_0)\mu_A\mu_Y}{n\pi_0}$$

When rss is the sampling design used for selecting the sample . The gain in accuracy due to the use of rss is measured by

$$G = V_{srs} - V_{(rss)} = \frac{(1-\pi_0)\sum_{i=1}^{m} \Delta^2_{Y_{(i)}}}{nm\pi_o} + \frac{(1-\pi_0)\left(\mu_Y^2 - \frac{\sum_{i=1}^{m} \mu^2_{Y_{(i)}}}{m}\right)}{n\pi_o} = G_1 + G_2$$

where

$$\sum_{i=1}^{m} \Delta^2_{Y_{(i)}} = \sum_{i=1}^{m} (\mu^2_{Y_{(i)}} + \mu_Y^2 - 2\mu_{Y_{(i)}}\mu_Y) = \sum_{i=1}^{m} \mu^2_{Y_{(i)}} - m\mu_Y^2$$

Then $G_1 = -G_2$ and the two designs are equivalent. This result reflects the fact that the ranking made by the expert does not provide an adequate ordering. That is, it is equivalent to the use of a random procedure. The particular structure of this optimization problem allows to determine easily an optimal set of probabilities by fixing intervals such that $\pi_h \in [a_h, b_h]$, $h=0,1,\ldots,k$. It is optimal, for a set of arbitrary selection probabilities, *that* $\sum_{h=1}^{k} \pi_h = 1-a_0$ Then the optimal strategy is to fix the value of $a_0 = \pi_0$.

3. ESTIMATION OF STIGMATIZED CHARACTERISTICS OF A HIDDEN GANG IN A FINITE POPULATION

Let s_1 and s_2 be independent samples selected by the srswr and denote their sizes by $\#s_j = n_j$, $j=1,2$. . Each interviewed selects between a sensitive question Q and an insensitive one Q^* using a random mechanism. It selects Q with probability π and Q^* with $1-\pi$. The persons should be convinced of the randomness of the selection of a number γ by the used mechanism. A certain value γ^* is fixed by the sampler. It determines a threshold and the interviewed has the stigma when $Y_i > \gamma^*$. He/she is considered as a member of a gang, say $G = \{u_j \in U \mid Y_i > \gamma\}$.

A random device $\Gamma(1)$ will be used by the individual in s_1. Each $u_j \in s_1$ performs the experiment and responds Y_i if he/she belongs to G. Else the response is the corresponding selected γ_{1t}. Nobody except the respondent knows the question that was answered. Then we may consider that the device generates γ with a known mean θ_1 and variance σ^2_1. The report is described by the model:

$$Z_{1i} = \begin{cases} Y_i & with \quad probability \quad \pi \\ \gamma_{1j} & with \quad probability \quad 1-\pi \end{cases}$$

A similar reasoning with the second sample s_2 with another random mechanism $\Gamma(2)$ associated to a known mean θ_2 and variance σ^2_2. Then we observe:

$$Z_{2i} = \begin{cases} Y_i & with \quad probability \quad \pi \\ \gamma_{2j} & with \quad probability \quad 1-\pi \end{cases}$$

The expectation under the model for $\Gamma(j)$ is $E(Z_{ij}) = \pi E(Y_j) = \pi E(Y) + (1-\pi)\theta_j = \pi\mu_Y + (1-\pi)\theta_j$ π is the unknown proportion of members of the gang .

$$\pi_s = 1 - \frac{\bar{z}_1 - \bar{z}_2}{\theta_1 - \theta_2} = 1 - \frac{d_{(srs)}}{D}$$

where

$$\bar{z}_{t\,t} = \frac{\sum_{i=1}^{n_j} Z_{ji}}{n_j}, \qquad j = 1,2$$

is an estimator of π when the sampling design is srswr .
The main result of the paper of Singh et.al. (1998) is:

Theorem 3.1. (Singh-Horn-Chowdury ,1998) π_s is unbiased and

$$V(\pi_s) = \frac{V_1 + V_2}{D^2} = \frac{V(d_{(srs)})}{D^2}$$

where

$$V^*_j = \frac{\pi\sigma_Y^2 + (1-\pi)\sigma_j^2 + (1-\pi)\pi(\mu_Y - \theta_j)^2}{n_j} = \frac{V_j}{n_j}, \qquad j = 1,2$$

An unbiased estimator of μ is given by

$$\mu_{srs} = \frac{\bar{z}_2\theta_1 - \bar{z}_1\theta_2}{D}$$

and

$$V(\mu_{srs}) = \frac{V_1(\theta_2 - \mu_Y)^2 + V_2(\theta_1 - \mu_Y)^2}{\pi^2 D^2}$$

We will develop a similar theorem when rss is the sampling design used for selecting the samples.

As usual we will assume that there is a cheap method for obtaining information for predicting *Y* for every sampled person u_i. Hence we are able to rank the selected individual without interviewing them. For example a look to the medical records of the selected persons permits to rank the possible level of their consumption of drugs *Y* using a concomitant variable *X*. Stokes (1977) considered the effect of the ranking errors due to the use of *X*. He obtained that it does not affect the main statistical properties of the rss mean estimator.

As is well known rss consists in the selection of m independent samples of size *m* using srswr. Generally m is not larger than 5. The individual in the sample are ranked. Take $Y_{(t:1)}, .., Y_{(t:t)}, \ldots Y_{(t:m)}$ as the order statistics (os) of the sample s_t. We measure only the os $Y_{(1::1)}, .., Y_{(t:t)}, \ldots Y_{(m:m)}$. The procedure is repeated r times (cycles). Denoting by $Y_{(t:t)k}$ the t-th os measured in the cycle *k* the rss mean is

$$\bar{y}_{rss} = \frac{\sum_{t=1}^{m}\sum_{k=1}^{r} Y_{(t:t)k}}{m^2 r} \tag{3.1}$$

It estimates unbiasedly μ_Y and

$$V(\bar{y}_{rss}) = \frac{\sum_{t=1}^{m}\sigma^2_{(t)}}{m^2 r} = \frac{\sigma^2}{mr} - \frac{\sum_{t=1}^{m}\left(\mu_{Y_{(t)}} - \mu_Y\right)^2}{m^2 r} = \frac{\sigma^2}{n} - \frac{\sum_{t=1}^{m}\Delta_{Y_{(t)}}{}^2}{mn} = \frac{\sigma^2}{n} - \frac{\Delta(Y)}{mn} \tag{3.2}$$

where $\mu_{(t)}$ and $\sigma^2_{(t)}$ denote the expectation and the variance of the os $Y_{(t:t)k}$ and $mr=n$.
The model of Singh et.al. (1998) considered the selection of two independent samples using srswr. We will select them using rss. The sample sizes are $n_j = r_j m_j$ and the use of the random mechanism generates responses

$$Z_j(t,k)=\begin{cases} Y_{j(t:t)k} & with \quad probability \quad \pi \\ \gamma_j(t,k) & with \quad probability \quad 1-\pi \end{cases} \qquad (3.3)$$

$$j=1,2, \quad ,t=1,...,m_j \quad k=1,...,r_j$$

Note that $\gamma_j(t,k)$ is a random variable and it is not ranked. Then the expectation of (3.3) under the procedure is:

$E[Z_j(t,k)]=\pi E(Y_{(t:t)k})+(1-\pi)E(\gamma_j(t,k))=\pi\mu_{Y(t)}+(1-\pi)\theta_j$

From Takahasi-Wakimoto (1988) we can derive that $\Sigma^{mj}_{t:=1}\Sigma^{rj}_{k=1}\mu_{Y(t)}=m_j r_j \mu_{Y.}$ Therefore the design expectation of the rss mean of (3.3) is:

$$E_d\left[\frac{\sum_{t=1}^{m_j}\sum_{k=1}^{r_j}Z_j(t,k)}{m_j r_j}\right]=E_d\left[\bar{z}_{j(rss)}\right]=\pi\mu_Y+(1-\pi)\theta_j=\mu_{j(rss)}$$

The variance of the rss estimator is easily derived from (3.2) as

$$V\left(\bar{z}_{j(rss)}\right)=\frac{\sum_{t=1}^{m_j}\sigma^2_{(t)}}{m_j{}^2 r_j}=\frac{\sigma^2}{m_j r_j}-\frac{\sum_{t=1}^{m_j}\left(\mu_{Y_{(t)}}-\mu_Y\right)^2}{m_j{}^2 r_j}=\frac{\sigma^2}{n_j}-\frac{\sum_{t=1}^{m}\Delta_{Y_{(t)}}{}^2}{m_j n_j}=\frac{\sigma^2}{n_j}-\frac{\Delta_j(Y)}{m_j\ n_j} \qquad (3.4)$$

$$j=1,2$$

The difference of the means $D(Z)=\mu_{1(rss)}-\mu_{2(rss)}$ can be expressed as $D(Z)=(1-\pi)(\theta_1-\theta_2)=(1-\pi)D$. Now we the proportion of persons in a gang is $\pi=1-D(Z)/D$ this is estimated by

$$\pi_{(rss)}=1-\frac{d_{(rss)}}{D}$$

where

$$d_{(rss)}=\frac{\sum_{t=1}^{m_1}\sum_{k=1}^{r_1}Z_1(t,k)}{m_1 r_1}-\frac{\sum_{t=1}^{m_2}\sum_{k=1}^{r_{12}}Z_2(t,k)}{m_2 r_2}=\bar{z}_{1(rss)}-\bar{z}_{2(rss)}$$

is the rss unbiased estimator of $D(Z)$.

The samples are independent then:

$$V\left(\pi_{(rss)}\right)=\frac{V_{1(rss)}+V_{2(rss)}}{D^2} \qquad (3.5)$$

Let us derive the $V_{j(rss)}$`s. Note that $Z_j(t,k)-E(Z_j(t,k))$ can take one of the two expressions

$S_1(t,k)=Y_{(t:t)k}-[\pi\mu_{Y(t)}+(1-\pi)\theta_j]=[Y_{(t:t)k}-\mu_{Y(t)}]-(1-\pi)[\theta_j-\mu_{Y(t)}]$

With probability π or

$S_2(t,k)]=\gamma_j(t,k)-[\pi\mu_{Y(t)}+(1-\pi)\theta_j)]=[\gamma_j(t,k)-\theta_j])-\pi[\mu_{Y(t)}-\theta_j].$

with probability $1-\pi$. Then

$(m_j r_j)V_{j(rss)}=\Sigma^{mj}_{t=1}\Sigma^{rj}_{k=1}[\pi E_d[S_1(t,k)]^2+(1-\pi)E_d[S_2(t,k)]^2] \qquad (3.6)$

Notice that for any j, t and k we have that $E_d[Y_{(t:t)k}-\mu_{Y(t)}][\theta_j-\mu_{Y(t)}]=E_d[\gamma_j(t,k)-\theta_j])][\mu_{Y(t)}-\theta_j]=0$,
$E_d[Y_{(t:t)k}-\mu_{Y(t)}]^2=\sigma^2_{(t)}$ is the variance of the t-th os of Y and $E_d[\gamma_j(t,k)-\theta_j]]^2=\sigma^2_j$ is the variance of γ for $\Gamma(j)$, $j=1,2$. Substituting these results in (3.6) we have

$$V_{j(rss)} = \frac{\sum_{t=1}^{m_j} \pi\sigma_{Y(t)}^2 + (1-\pi)\sigma_j^2 + \pi(1-\pi)(\theta_j - \mu_{Y(t)})^2}{m_j r_j^2} \qquad j = 1,2 \tag{3.7}$$

using the relation between the variance of an os and the variance of the variable, we rewrite this expression as:

$$V_{j(rss)} = \frac{\pi\sigma^2 + (1-\pi)\sigma_j^2}{n_j} + \frac{\pi(1-\pi)\sum_{t=1}^{m_j}(\theta_j - \mu_{Y(t)})^2}{m_j n_j} - \left[\frac{\pi\sum_{t=1}^{m_j}(\mu_{Y(t)} - \mu_Y)^2}{m_j n_j}\right] \quad j = 1,2 \tag{3.8}$$

Compare (3.8) with the variance derived by Singh et. al. (1998) and note that the last term is the gain in accuracy due to the use of rss.

The estimation $\pi_{(rss)}$ is plugged-in and plugging-in the estimator we derive the variable

$$Y^*_{j(t:t)k} = \frac{Z_j(t,K) + (1-\pi_{(rss)})\theta_j}{\pi_{(rss)}}$$

The estimator derived by using the corresponding rss estimators is :

$$\mu_{(rss)} = \frac{\sum_{t=1}^{m_j}\sum_{k=1}^{r_j} Y^*_{j(t:t)k}}{m_j r_j} = \frac{\theta_1 \bar{z}_{2(rss)} - \theta_2 \bar{z}_{1(rss)}}{(\bar{z}_{2(rss)} - \theta_2)(\bar{z}_{1(rss)} - \theta_1)} = \frac{d_1}{d_2} \tag{3.9}$$

Note that it has the same structure that $\mu_{srs.}$ Then we have that

$$V(\mu_{(rss)}) = \frac{V_{1(rss)}(\theta_2 - \mu_Y)^2 + V_{2(rss)}(\theta_1 - \mu_Y)^2}{\pi^2 D^2} \tag{3.10}$$

Now, we can establish an rss counterpart of Theorem 1.

Theorem 3.2. When the procedure proposed by Singh et. al. (1998) is used for obtaining the randomized responses and the sampling design is rss we have:

1. The estimator of the proportion of persons in the gang $\pi_{(rss)} = 1 - \frac{d_{(rss)}}{D}$ is unbiased with variance (3.5) and its gain in accuracy is

$$\frac{\pi}{D^2}\left[\sum_{j=1}^{2} \frac{\sum_{t=1}^{m_j}(\mu_{Y(t)} - \mu_Y)^2}{m_j n_j}\right]$$

2. $\mu_{(rss)}$ is an unbiased estimator of μ_Y with variance (3.10) and its gain in accuracy is

$$\frac{1}{\pi D^2}\left[\frac{\sum_{t=1}^{m_1}(\mu_{Y(t)} - \mu_Y)^2}{m_j n_j}(\theta_2 - \mu_Y)^2 + \frac{\sum_{t=1}^{m_2}(\mu_{Y(t)} - \mu_Y)^2}{m_j n_j}(\theta_1 - \mu_Y)^2\right]$$

Then the surveyor should manage to fix π and D for ensuring a large gain in accuracy

RECEIVED OCTOBER 2007
REVISED DECEMBER 2008

ACKNOWLEDGEMENTS:
This paper is a result obtained within the project Modelos Matemáticos para la Toma de Decisiones en Sistemas Complejos bajo Incertidumbre, supported by the Ministerio de Ciencias Tecnología y Medio Ambiente, Cuba

REFERENCES

[1] ARNOLD, B.C. N. BALAKRISHNAN and H.N. NAGARAYA (1992): **A First Course in Order Statistics**. Wiley, N. York.

[2] BARABASI, L. and C. PISANI (2002): Ranked set sampling for replicated sampling designs. **Biometrics.** 58, 586-592.

[3] BOUZA, C.N. (2002a): Estimation of the mean in ranked set sampling with non responses **Metrika,** 56, 171 – 179

[4] BOUZA, C.N. (2002b):. Ranked set subsampling the non response strata for estimating the difference of means. Biometrical. J. 44, 903-915.

[5] CHAUDHURI, A., T. MAITI and R. ROY (1996): A note on competing variance estimators in randomized response surveys. **Austral. and N. Zealand J. Statistic**. 38, 35-4.

[6] CHAUDHURY A. and H. STENGER (2005): **Sampling Survey**. M. Dekker, N. York.

[7] MCINTYRE, G.A. (1952): A method of unbiased selective sampling using ranked sets. **J. Agric. Res**. 3, 385-390.

[8] PATIL GP, SINHA A.K and TAILLIE C (1994) Ranked set sampling. In: Patil GP, Rao CR (eds) **Handbook of Statistics**, vol 12. North Holland, Amsterdam

[9] PATIL GP, SINHA A.K and TAILLIE C (1999) Ranked set sampling: a bibliography. **Environ Ecol Stat** 6:91–98

[10] PATIL G.P (2002) **Ranked set sampling**. In: El- Shaarawi AH, Pieegoshed WW (eds) Encyclopedia of enviromentrics, vol 3. Wiley, Chichester pp 1684–1690

[11] SINGH S. , S. HORN and H. CHOWDHURY (1998): Estimation of stigmatized characteristics of a hidden gang in a finite population. **Austral. And N. Zealand J. of Stat**. 40, 291-297.

[12] SINGH S, SING R (1993) Generalized Franklin's model for randomized response sampling. **Commun Stat ATheory Methods** 22:741–755

[13] TAKAHASI K. and K. WAKIMOTO (1968): On unbiased estimates of population mean based on the sample stratified by means of ordering. **Ann. Intern. Mathem. Stat**. 20, 1-31.

[14] WARNER, S. L. (1965): Randomized responses: a survey technique for eliminating evasive survey bias. **J. American Statistic. Ass.** . 60, 63-69.

[15] ZOU, G. (1997): Two-stage randomized response technique as a single stage procedure. **Austral. and N. Zeeland J. Stat.**39, 235-236