RANKED SET SAMPLING AND OPTIONAL SCRAMBLING RANDOMIZED RESPONSE MODELING

Carlos N. Bouza[1]*, Poonam Singh**[2] , Rajesh Singh**, José F. García***
* Facultad de Matemática y Computación, Universidad de La Habana,
**Department of statistics, Banaras Hindu University (U.P), India
***Universidad Juárez Autónoma de Tabasco, México

ABSTRACT
Ranked Set Sampling (RSS) strategies when an Optional in Scrambling Randomized Response Modeling procedure is used, are derived. In this paper are obtained Ranked Set Sampling models for estimating unbiasedly the mean, the variance of the sensitive variable as well as of its sensitivity level W. The scrambled response modeling of Mehta et al (2012) or Huang (2010) are used.

KEYWORDS: Ranked Set Sampling, Randomized Response, Scrambled Variable, Unbiased Estimations

MSC: 62D05

RESUMEN
Estrategias Ranked Set Sampling (RSS) cuando es usado el llamado procedimiento “Optional in Scrambling Randomized Response Modeling”**,** son derivadas. En este paper se obtienen modelos RSS para estimar insesgadamente la media, la varianza de la variable sensitiva, así como del nivel de sensibilidad W. El modelo de enmascaramiento de Mehta et al (2012) o Huang (2010) son usados.

PALABRAS CLAVE: Ranked Set Sampling, Respuestas Aleatorizadas, Variable Enmascarada, Estimaciones Insesgadas

1. INTRODUCTION

Commonly, the development of survey sampling inquiries deals with the existence of missing data or response biases, see an interesting discussion in Rueda and Gonzalez (2004). A particularly complicated problem appears when is required to obtain information on sensitive attributes. Collecting trustworthy responses on sensitive issues, through direct questioning in personal interviews, often is unsuccessful because is unprotected the respondents’ privacy. Therefore, in practice the data collected on sensitive features are affected by the existence of respondent bias.

Randomized response (RR) models are used to decrease both non-response and answer bias and to provide privacy protection to the respondents. Warner (1965) proposed the so called randomized response (RR) method as a mean of avoiding response bias. The initial model looked for the estimation of the proportion of persons with a stigma. The model used a randomized trial, for selecting between responding the sensitive question or an insensitive one. This technique protects the privacy of the respondent by granting that belonging to a stigmatized group cannot be detected. After the seminal paper of Warner (1965), different contributions have been proposed and they generated a large set of models. RR provides the opportunity of reducing response biases due to dishonest answers to sensitive questioning.

Consider a population U of size N with two strata U_A and U_{A^*} where belonging to U_A is stigmatizing. The interest of the inquiry is to estimate the proportion of individuals carrying a stigma, identified with the belonging to A. If $|A|$ denotes the number of units with the stigma and we are interested in estimating the probability $W = |U_A| / |U| = N_A/N$.

The RR technique has been successfully applied in many areas and different modifications and extensions on this method have proposed the literature on sampling, see Chaudhuri and Mukerjee (1988). It is still receiving

---
[1] bouza@matcom.uh.cu
[2] ppiyu123@gmail.com

attention from the researchers, see for example Bar et al (2004), Bouza (2010), Chistofides (2003), Gupta et al (2012), Huang (2010) Mehta et al (2012) and Ryu et al.. (2005),

The RR procedures are generally based on the selection of a sample using simple random sampling with replacement (SRSWR). We will consider the use of ranked set sampling (RSS) for selecting the sample. It is an alternative sample design which generally provides gains in accuracy with respect to SRSWR. McIntyre proposed it for estimating pastures yield in 1952, see Bouza (2013), Chen et al. (2004) and Al-Omari et al. (2008) for details. He established that to estimate the mean pasture yield using RSS was more efficient than to select the sample using a SRSWR design. The RSS considers that the units may be ranked by means of a cheap procedure. Then an order statistics is selected from each of the independent samples selected using SRSWR. It turned out that the use of RSS is highly beneficial and leads to estimators which are more precise than the usual sample mean per unit ones.

McIntire (1952) proposed the method of ranked set sampling (RSS). He proposed that the units may be ranked visually and noticed the existence of a gain in accuracy with respect to the use of the sample mean when SRSWR was used. Dell and Clutter (1972) and Takahasi and Wakimoto (1968) provided mathematical support to his claims. A practical problem is to obtain an accurate and cheap ranking procedure. If it is not accurate we will have errors in the rankings. Dell-Clutter (1972) established that in such case we obtain not the i^{th} -os but the 'i^{th} – judgmental' one. The use of a concomitant variable is supported by the existence of a correlation between the true variable and the ranking variable. Stokes (1977) and Patil et al. (1995a) studied the use of concomitant variables and used them for ranking.

Bouza (2009) developed a RSS strategy when a RR procedure is used for obtaining information on a sensitive variable. In this paper we consider that a respondent can report the truth or scramble the response when the study variable is sensitive. We are also interested in the proportion of respondents reporting the scrambled response (sensitivity level of the study variable). In this paper we develop RSS models for the unbiased estimation of the mean, the variance and the sensitivity level W, of a sensitive variable, when the scrambled response modeling of Mehta et al (2012) or Huang (2010) are used. The ranking using a related variable to Y is analyzed as well as the ranking using the scrambled variables, for applying the RR procedures. We derive that the use of RSS produces considerable gains in accuracy and provide a measurement of it.

2. THE RR PROCEDURES OF HUANG ET AL. AND OF MEHTA ET AL. UNDER SRSWR

The RR proposed by Huang et al (2010) is to use a pair of randomization mechanisms for collecting responses in each of the independent samples of size n_i (i=1,2), drawn from using SRSWR . The respondent $j \in s_i$ selects randomly between reporting the true response Y.

For reporting, the interviewee performs, independently, experiments of Bernoulli, for each i, j such that $E(S_{ij})=1$, $V(S_{ij})= \vartheta_i^2$; $E(C_j)=W$ and generates randomly a variable D_{ij} of distribution is fixed by the sampler with $E(D_{ij}) = \bar{D}_i$ and $V(D_{ij}) = \delta_i^2$; $\bar{D}_1 \neq \bar{D}_2$. The report of the jth interviewee of the sample s_i is modeled by

$$Z_{ij}^* = (1 - C_j)Y_j + C_j(S_{ij}Y_j + D_{ij})$$

The expected value and variance of the report are

$$E(Z_{ij}^*) = (1 - W)\mu_Y + W(\mu_Y + \bar{D}_i)$$

$$V_{Hi}^2 = V(Z_{ij}^*) = \sigma_Y^2 + W(\mu_Y^2 + \sigma_Y^2)\vartheta_i^2 + W(1 - W)\bar{D}_i^2 + W\delta_i^2$$

As $\bar{D}_i = E(D_{ij})$ is known an unbiased estimator of the mean of Y is easily derived . It is

$$\hat{\mu}_H = \frac{\bar{D}_1\bar{Z}_2^* - \bar{D}_2\bar{Z}_1^*}{\bar{D}_1 - \bar{D}_2}, \bar{Z}_i^* = \frac{\sum_{j=1}^{n_i} Z_{ij}^*}{n_i}, i = 1,2$$

Its variance is given by

$$V(\hat{\mu}_H) = \frac{\bar{D}_1^2 V_{H2}^2}{n_2(\bar{D}_1 - \bar{D}_2)^2} + \frac{\bar{D}_2^2 V_{H1}^2}{n_1(\bar{D}_1 - \bar{D}_2)^2},$$

This method has been studied in different papers as Hussain Z, Al-Sobhi MM, Al-Zahrani B (2014)

3. RSS MODELS

3.1 Basic ideas on RSS

McIntire (1952) proposed the method of ranked set sampling (RSS). He used the fact that the units may be ranked cheaply, and noticed the existence of a gain in accuracy, with respect to the use of the sample mean, when simple random sampling with replacement (SRSWR) was used. Dell and Clutter (1972) and Takahasi and Wakimoto (1968) provided mathematical support to his claims. The following algorithm provides a description of ranked set sampling [RSS].

RSS implementation

```
Input r, m
j=0 and t=0
While t<r+1 do
While j<m+1 do
Select a sample s_j of size |s_j|=m_j using SRSWR.
Rank the sampled units with respect to the variable of interest ξ
Measure the variable in the unit with rank i (ξ_{j:m)t})
j=j+1
End
t=t+1
End
```

The RSS sample is the sequence of order statistics (os) $\xi_{(1:1)t},\ldots, \xi_{(m:m)t}$, where $(j:h)_t$ denotes the statistic of order j in the h^{th} sample in the cycle $t=1,..,r$. We have $n=mr$ observation and r of them are of the *i-th* order statistics (os), $i=1,...,m$. The RSS estimator of the mean of a variable of interest ξ , μ_ξ is

$$\mu_{(rss)\xi}=\sum^{r}_{t=1} \sum^{m}_{i=1}\xi_{(i:m)r}/rm$$

and its variance is given by

$$V(\mu_{(rss)\xi})= \sum^{m}_{i=1} \sigma^2_{\xi(i:m)}/rm^2=[\sigma^2_{\xi}-\sum^{m}_{i=1} \Delta^2_{(i:m)}]/rm$$

where

$$\sigma^2_{\xi(i:m)} =E[\xi_{(i:m)} - E([\xi_{(i:m)})]^2$$

and

$$\Delta_{(i:m)}= E([\xi_{(i:m)})]-_I \mu_\xi$$

3.2 RSS for the RR of Huang

Let us develop the RSS model for the RR procedure proposed by Huang et al (2010). The sampler ranks the units selected using some available information on Y. As in the SRSWR case we assume that the respondent $j\in s_{ijt}$, $i=1,2; j=1,\ldots,m_i; t=1,..,r$ selects randomly between reporting the true response Y or a scrambled response. Now in each ordered sample s_{ij} is interviewed the person with the jth-rank. $Y_{i(j)t}, j = 1, \ldots, m_i$. We repeat the selection in each cycle $t=1,\ldots,r$. We will take for simplifying the notation m_i=m, i=1, 2; hence $n_i=mr$.
The sampler ranks the individual in each sample using auxiliary information on Y. and evaluates the j^{th}-uit in the sample s_{ijt} . The report is modeled by:

$$Z^*_{i(j)t} = \left(1 - C_{ijt}\right)Y_{i(j)t} + C_{ijt}\left(S_{ijt}Y_{i(j)t} + D_{ijt}\right); i = 1{,}2; j = 1,..,m; t = 1,..,r$$

The expected value and variance of each report are

$$E(Z^*_{i(j)t}) = (1 - W)\mu_{Y_{(j)}} + W(\mu_{Y_{(j)}} + \bar{D}_i)$$

$$V_{Hi(j)t}^2 = V(Z_{i(j)t}^{*}) = \sigma_{Y_{(j)}}^2 + W\left(\mu_{Y_{(j)}}^2 + \sigma_{Y_{(j)}}^2\right)\vartheta_i^2 + W(1-W)\bar{D}_i^2 + W\delta_i^2$$
$$= \sigma_Y^2 - \Delta_{(j)}^* + W\left(\mu_{Y_{(j)}}^2 + \sigma_Y^2 - \Delta_{(i)}^*\right)\vartheta_i^2 + W(1-W)\bar{D}_i^2 + W\delta_i^2$$
$$= \sigma_Y^2\left(1 + W\vartheta_i^2\right) + W\vartheta_i^2\mu_{Y_{(j)}}^2 + W(1-W)\bar{D}_i^2 + W\delta_i^2 - \Delta_{(j)}^*\left(1 + W\vartheta_i^2\right),\ \Delta_{(j)}^*$$
$$= \left(\mu_{Y_{(j)}} - \mu_Y\right)^2, i = 1,..,m\ ; j = 1,..,m; t = 1,..r$$

Take

$$\hat{\mu}_{HRSSi} = \frac{\sum_{t=1}^{r}\sum_{j=1}^{m} Z_{i(j)t}^*}{mr} = \frac{\sum_{t=1}^{r}\sum_{i=1}^{m}(1-C_{ijt})Y_{i(j)t}}{mr} + \frac{\sum_{t=1}^{r}\sum_{i=1}^{m} C_{iit}\left(S_{ijt}Y_{i(j)t} + D_{ijt}\right)}{mr}, i = 1,2; j = 1, \dots, m$$

Its expectation is

$$E(\hat{\mu}_{HRSSi}) = \frac{(1-W)\sum_{i=1}^{m}\mu_{Y_{(i)}}}{m} + \frac{W\sum_{i=1}^{m}\mu_{Y_{(i)}}}{m} + W\bar{D}_i = \mu_Y + W\bar{D}_i, i = 1,2$$

Then, we have the unbiasedness of

$$\hat{\mu}_{HRSS} = \frac{\bar{D}_1\hat{\mu}_{HRSS2} - \bar{D}_2\hat{\mu}_{HRSS1}}{\bar{D}_1 - \bar{D}_2}$$

Due to the independence

$$V(\hat{\mu}_{HRSS}) = \frac{\bar{D}_1^2 V(\hat{\mu}_{HRSS2}) + \bar{D}_2^2 V(\hat{\mu}_{HRSS1})}{(\bar{D}_1 - \bar{D}_2)^2}$$

where

$$V(\hat{\mu}_{HRSSi}) = \frac{\sigma_Y^2\left(1 + W\vartheta_i^2\right)}{n_i} - \frac{\left(1 + W\vartheta_i^2\right)\sum_{j=1}^{m}\Delta_{(j)}^*}{n_i m} + \frac{W\vartheta_i^2\sum_{j=1}^{m}\mu_{Y_{(j)}}^2}{n_i m} + \frac{W(1-W)\bar{D}_i^2 + W\delta_i^2}{n_i},\ \Delta_{(j)}^* = \left(\mu_{Y_{(j)}} - \mu_Y\right)^2, i = 1,2; j = 1, \dots m$$

Comparing this variance with the corresponding to the SRSWR model with $n_i = \frac{n}{2}$ we have

$$V(\hat{\mu}_{HRSSi}) - V_{Hi}^2 = W\vartheta_i^2\left(\frac{\sum_{j=1}^{m}\mu_{Y_{(j)}}^2}{nm/2} - \mu_Y^2\right) - \frac{(1+W\vartheta_i^2)\sum_{i=1}^{m_i}\Delta_{(j)}^*}{nm/2} = -\frac{\sum_{j=1}^{m_i}\Delta_{(j)}^*}{nm/2}, \text{ i=1,2}$$

Hence, the gain in accuracy due to using a RSS sample, ranking with respect to Y is

$$G_H(SRSWR, RSS) = V(\hat{\mu}_{HRSS}) - V(\hat{\mu}_H) = \sum_{i=1}^{2}\frac{\bar{D}_i^2}{(\bar{D}_1 - \bar{D}_2)^2}\frac{\sum_{j=1}^{m_i}\Delta_{(i)}^*}{nm/2}$$

As it is positive RSS should be preferred.

In the RSS case the sensitive is estimated unbiasedly by

$$\hat{\mu}_{HRSS}^W = \frac{\hat{\mu}_{HRSS2} - \hat{\mu}_{HRSS1}}{\bar{D}_1 - \bar{D}_2}$$

Its variance is

$$V(\hat{\mu}_H^W) = \frac{V(\hat{\mu}_{HRSS2})}{(\bar{D}_1 - \bar{D}_2)^2} + \frac{V(\hat{\mu}_{HRSS1})}{(\bar{D}_1 - \bar{D}_2)^2}$$

From the above given results we have tat the gain in accuracy of the proposed estimator is

$$G_H^W(SRSWR, RSS) = \frac{1}{(\bar{D}_1 - \bar{D}_2)^2}\sum_{i!}^{2}\frac{\sum_{j=1}^{m_i}\Delta_{(j)}^*}{nm/2} =$$

4. USING THE SCRAMBLING VARIABLES FOR RANKING.

The scrambling variables are independent of the variable of interest. Let us see what happens when the ranking variable is a scrambling variable. That is, when the rank is made using the variables introduced by the RR procedure.

The inquiry may be developed as follows. The respondents generate the scrambling variable and they communicate their values. Then they fix their rank in the sample without informing of the value to the sampler. Now ξ is the scrambled variable which distribution is completely known.

Let us consider the procedure of Huang et al (2012) and the ranking using ξ=S. Now the report is modeled by

$$Z_{i(j)t}^{S}=(1-C_{ijt})Y_{ijt}+C_{ijt}\left(S_{i(j)t}Y_{ijt}+D_{ijt}\right)$$

The expected value of the report is the expectation of the j[th] statistic of order of the scrambled variable S is: $E(S_{i(j)t})=\mu_{i(j)}^{S}$. Hence

$$E(Z_{i(j)t}^{S})=(1-W)\mu_{Y}\ \ +W(\mu_{i(j)}^{S}\mu_{Y}\ \ +\overline{D}_{i})$$

The distribution of S is fixed by the sampler and as a consequence the parameters of it are known. The variance of the report is given by

$$Z_{i(j)t}^{S}=(1-C_{ijt})Y_{ijt}+C_{ijt}\left(S_{i(j)t}Y_{ijt}+D_{ijt}\right)$$

$$\begin{aligned}V_{i(j)}^{S2}=V(Z_{i(j)t}^{S}\ \ )&=\sigma_{Y}^{2}\ \ +W\left(\mu_{Y}^{2}\ \ +\sigma_{Y}^{2}\ \ \right)\left(\vartheta_{i}^{2}-\Delta_{i(j)}^{S}\right)+W(1-W)\overline{D}_{i}^{2}+W\delta_{i}^{2}\\&=\sigma_{Y}^{2}\ \ +W\left(\mu_{Y}^{2}\ \ +\sigma_{Y}^{2}\ \ \right)\vartheta_{i}^{2}+W(1-W)\overline{D}_{i}^{2}+W\delta_{i}^{2}-W\left(\mu_{Y}^{2}\ \ +\sigma_{Y}^{2}\ \ \right)\Delta_{i(j)}^{S};\ \Delta_{i(j)}^{S}\\&=\left(\mu_{i(j)}^{S}-1\right)^{2},\end{aligned}$$

Then we have that

$$\hat{\mu}_{HRSSi}^{S}\ \ =\frac{\sum_{t=1}^{r}\ \ \sum_{j=1}^{m}Z_{i(j)t}^{S}}{mr}$$

has expectation

$$E(\hat{\mu}_{HRSSi}^{S})=(1-W)\mu_{Y}\ \ +W\mu_{Y}\ \ +W\overline{D}_{i}=\mu_{Y}+W\overline{D}_{i},i=1,2$$

As a result, we have the unbiasedness of the estimator

$$\hat{\mu}_{HRSS}^{S}=\frac{\overline{D}_{1}\hat{\mu}_{HRSS2}^{S}-\overline{D}_{2}\hat{\mu}_{HRSS1}^{S}}{\overline{D}_{1}-\overline{D}_{2}}$$

Its sampling error is

$$V(\hat{\mu}_{HRSS}^{S})\ \ =\frac{\overline{D}_{1}^{2}V_{2}^{S2}+\overline{D}_{2}^{2}V_{1}^{S2}}{(\overline{D}_{1}-\overline{D}_{2})^{2}},$$

where

$$V_{i}^{S2}=\frac{\sigma_{Y}^{2}}{n_{i}}+\frac{W\left(\mu_{Y}^{2}\ \ +\sigma_{Y}^{2}\ \ \right)\vartheta_{i}^{2}+W(1-W)\overline{D}_{i}^{2}+W\delta_{i}^{2}}{n_{i}}-\frac{W\left(\mu_{Y}^{2}\ \ +\sigma_{Y}^{2}\ \ \right)\sum_{j=1}^{m}\ \ \Delta_{i(j)}^{S}}{n_{i}m},i=1,2$$

Now we have that the diminution in the variance is measured by

$$G_{H}^{S}(SRSWR,RSS)=\frac{W\left(\mu_{Y}^{2}\ \ +\sigma_{Y}^{2}\ \right)}{(\overline{D}_{1}-\overline{D}_{2})^{2}}\sum_{i=1}^{2}\frac{\overline{D}_{i}^{2}}{nm/2}\sum_{j=1}^{m}\ \ \Delta_{i(j)}^{S}$$

It is positive and of course the larger the better is the use of RSS.

Let us compare this ranked model estimation with using RSS ranking with respect to Y. Taking

$$\begin{aligned}I\left\{G_{H}\left(SRSWR,RSS(Y)\right)vsG_{H}^{S}(SRSWR,RSS)\right\}&=G_{H}\left(SRSWR,RSS(Y)\right)-G_{H}^{S}(SRSWR,RSS)\\&=\sum_{i=1}^{2}\frac{\overline{D}_{i}^{2}}{(\overline{D}_{1}-\overline{D}_{2})^{2}}\frac{\sum_{j=1}^{m}\Delta_{i(j)}^{*}}{\frac{nm}{2}}-W\left(\mu_{Y}^{2}\ \ +\sigma_{Y}^{2}\ \right)\sum_{i=1}^{2}\frac{\overline{D}_{i}^{2}}{(\overline{D}_{1}-\overline{D}_{2})^{2}}\frac{\sum_{j=1}^{m}\ \ \Delta_{i(j)}^{S}}{\frac{nm}{2}}\\&=\sum_{i=1}^{2}\frac{\overline{D}_{i}^{2}}{(\overline{D}_{1}-\overline{D}_{2})^{2}}\frac{\sum_{j=1}^{m}\Delta_{i(j)}^{*}-W\left(\mu_{Y}^{2}\ \ +\sigma_{Y}^{2}\ \right)\Delta_{i(j)}^{S}}{\frac{nm}{2}}\end{aligned}$$

Therefore is better to rank using S if $I\left\{G_{H}\left(SRSWR,RSS(Y)\right)vsG_{H}^{S}(SRSWR,RSS)\right\}<0$.

The proposed estimator of the sensitive is

$$\hat{\mu}_{HRSS}^{WS}=\frac{\hat{\mu}_{HRSS2}^{S}\ \ -\hat{\mu}_{HRSS1}^{S}}{\overline{D}_{1}-\overline{D}_{2}}$$

and

$$V(\hat{\mu}_{HRSS}^{WS})=\frac{V(\hat{\mu}_{HRSS2}^{S})}{(\overline{D}_{1}-\overline{D}_{2})^{2}}+\frac{V(\hat{\mu}_{HRSS1}^{S})}{(\overline{D}_{1}-\overline{D}_{2})^{2}}$$

From the above given results we have that the gain in accuracy of the proposed estimator is

$$G_H^{WS}(SRSWR, RSS) = \frac{W\left(\mu_Y^2 \quad + \sigma_Y^2 \ \right)}{(\bar{D}_1 - \bar{D}_2)^2} \sum\nolimits_{i=1}^{2} \frac{1}{n_i m_i} \sum\nolimits_{j=1}^{m_i} \quad \Delta_{i(j)}^S$$

The increase in precision is easily derived form the result obtained for the estimation of the mean of Y.
If we use for ranking ξ=D instead of S we have:

$$Z_{i(j)t}^D = (1 - C_{ijt})Y_{ijt} + C_{ijt}\left(S_{ijt}Y_{ijt} + D_{i(j)t}\right)$$

The expectation and variance of this report are

$$E(Z_{i(j)t}^D) = (1 - W)\mu_Y \quad + W(\mu_Y \quad + \mu_{(j)}^D)$$

and

$$V_{i(j)}^{D2} = V(Z_{i(j)t}^D \quad) = \sigma_Y^2 \quad + W\left(\mu_Y^2 \quad + \sigma_Y^2 \ \right)\vartheta_i^2 + W(1 - W)\mu_{i(j)}^{D2} + \quad W\left(\delta_i^{D2} - \Delta_{i(j)}^D\right);\ \Delta_{i(j)}^D = \left(\mu_{i(j)}^D - \bar{D}_i\right)^2,$$
$$j = 1, .., m \quad ; i = 1,2$$

We propose as RSS estimator, using the ranking in the scrambled variable D in the subsample s_i,

$$\hat{\mu}_{HRSSi}^D \quad = \frac{\sum_{t=1}^r \quad \sum_{j=1}^m Z_{i(j)t}^D}{mr}$$

Its expected value is

$$E(\hat{\mu}_{HRSSi}^D) = (1 - W)\mu_Y \quad + W\mu_Y \quad + W\bar{D}_i = \mu_Y + W\bar{D}_i, i = 1,2$$

Then, we have the unbiasedness of

$$\hat{\mu}_{HRSS}^D = \frac{\bar{D}_1\hat{\mu}_{HRSS2}^D - \bar{D}_2\hat{\mu}_{HRSS1}^D}{\bar{D}_1 - \bar{D}_2}$$

Its sampling error is

$$V(\hat{\mu}_{HRSS}^D) \quad = \frac{\bar{D}_1^2 V_2^{S2} + \bar{D}_2^2 V_1^{S2}}{(\bar{D}_1 - \bar{D}_2)^2},$$

where

$$V_{i(j)}^{D2} = V(Z_{i(j)t}^D \quad) = \sigma_Y^2 \quad + W\left(\mu_Y^2 \quad + \sigma_Y^2 \ \right)\vartheta_i^2 + W(1 - W)\mu_{i(j)}^{D2} + \quad W\left(\delta_i^{D2} - \Delta_{i(j)}^D\right);\ \Delta_{i(j)}^D = \left(\mu_{i(j)}^D - \bar{D}_i\right)^2,$$
$$j = 1, .., m \quad ; i = 1,2$$

$$V_i^{D2} = \frac{\sigma_Y^2}{n_i} + \frac{W\left(\mu_Y^2 \quad + \sigma_Y^2 \ \right)\vartheta_i^2 \quad + W\delta_i^2}{n_i} + \frac{W(1 - W)\sum_{j=1}^m \mu_{i(j)}^{D2} - W\sum_{j=1}^{m_i} \quad \Delta_{i(j)}^D}{n_i m_i}, i = 1,2$$

Now we have that the increase in the precision due to RSS with respect to SRSWR is

$$G_H^D(SRSWR, RSS) = \frac{W}{(\bar{D}_1 - \bar{D}_2)^2 nm/2} \sum\nolimits_{i=1}^{2} \quad \left[\sum\nolimits_{j=1}^{m} \Delta_{i(j)}^2 - (1 - W)\left(\bar{D}_i^2 - \sum_{j=1}^{m} \mu_{i(j)}^{D2}\right)\right]$$

Comparing this design with ranking with respect to Y, the increase due to using a ranking using D is

$$I\{G_H\left(SRSWR, RSS(Y)\right) vs G_H^D(SRSWR, RSS)\} = G_H\left(SRSWR, RSS(Y)\right) - G_H^D(SRSWR, RSS) =$$
$$\frac{W}{(\bar{D}_1 - \bar{D}_2)^2 nm/2}\left[\sum\nolimits_{i=1}^{2}\left(\mu_Y^2 \quad + \sigma_Y^2 \ \right)\sum\nolimits_{j=1}^{m} \quad (\Delta_{i(j)}^S - \Delta_{i(j)}^2) + \sum\nolimits_{i=1}^{2} \quad \left[\sum\nolimits_{j=1}^{m} \quad (1 - W)\left(\bar{D}_i^2 - \sum_{j=1}^{m} \mu_{i(j)}^{D2}\right)\right]\right]$$

For estimating the sensitive is proposed

$$\hat{\mu}_{HRSS}^{WD} = \frac{\hat{\mu}_{HRSS2}^D \quad - \hat{\mu}_{HRSS1}^D}{\bar{D}_1 - \bar{D}_2}$$

as

$$V(\hat{\mu}_{HRSS}^{WS}) = \frac{V(\hat{\mu}_{HRSS2}^{D})}{(\bar{D}_1 - \bar{D}_2)^2} + \frac{V(\hat{\mu}_{HRSS1}^{D})}{(\bar{D}_1 - \bar{D}_2)^2}$$

The analysis of the increase in accuracy of using D as the ranked variable in estimating W, is derived directly from the results obtained for $\hat{\mu}_{HRSS}^{D}$.

Let us compare the Scrambled RSS estimators.

The relation

$$I\{G_H^S(SRSWR, RSS(Y)) vs G_H^D(SRSWR, RSS)\} = G_H^S(SRSWR, RSS) - G_H^D(SRSWR, RSS) G_H\ (RSS(S), RSS)$$

$$= \sum_{i=1}^{2} \frac{W\bar{D}_i^2}{n_i m_i (\bar{D}_1 - \bar{D}_2)^2} \left((\mu_Y^2 \ \ + \sigma_Y^2\ ) \sum_{j=1}^{m_i} \ \ \Delta_{i(j)}^S - W \sum_{j=1}^{m_i} \Delta_{i(j)}^D \ \right)$$

measures the increase in accuracy of using S with respect to D.

Apparently the ranking in S generally provides a larger gain in accuracy. This result allows developing easily that the measure of the gain is accuracy for the sensitivity is

$$I\{G_H^{WS}(SRSWR, RSS(Y)) vs G_H^{WD}(SRSWR, RSS)\} =$$

$$= \sum_{i=1}^{2} \frac{W}{n_i m_i (\bar{D}_1 - \bar{D}_2)^2} \left((\mu_Y^2 \ \ + \sigma_Y^2\ ) \sum_{j=1}^{m_i} \ \ \Delta_{i(j)}^S - W \sum_{j=1}^{m_i} \Delta_{i(j)}^D \ \right)$$

Note that the sampler has the possibility of tuning the accuracy by fixing the distribution of S and D conveniently.

5. CONCLUSIONS

Ranking using a scrambling variable is not subject to error as is general the case when an auxiliary variable is used for ranking the interest variable *Y*.

The un-knowledge of the behavior of the sensitive variable *Y* is not an impediment for using RSS when one of the studied RR procedures is used.

The accuracy related with ranking, using a scrambled variable of the RR procedure, seems to be generally larger than the obtained ranking *Y*.

The sampler is able to control the diminution of the variance of the estimators by tuning conveniently the parameters of the distribution of the ranking scrambling variable.

Acknowledgements: One of the authors was supported by the project BigData Agross. The authors thank the referees for their helpful comments.

RECEIVED: FEBRUARY, 2017.
REVISED: MAY, 2017.

REFERENCES

[1] AL-OMARI, A. I. and K. H. JABER (2008): Percentile double ranked set sampling, **J. Math. Stat**. 4 60-64.

[2] BAR-LEV S. K, BOBOVITCH E., and BOUKAI, B. (2004): A note on randomized response models for quantitative data. **Metrika** 60, 255–260.

[3] BOUZA C. N. (2013): **Handling with Missing Data in Ranked Set Sampling**. Springer Briefs, Springer Verlag, Germany.

[4] BOUZA C. N. (2009): Ranked set sampling and randomized response procedures for estimating the mean of a sensitive quantitative character. **Metrika** 70, 267-277.

[5] CHAUDHURi, A. and R. MUKERJEE (1988): **Randomized Response Theory and Techniques**, Marcel Dekker, New-York, 1988.

[6] CHEN, Z., Z. BAI and K. B. SINHA (2004**): Ranked Set Sampling.** Springer, Berlin.

[7] CHRISTOFIDES, T.C. (2003): A generalized response technique. **Metrik**a. 57, 195-200.

[8] DELL, J.R. and CLUTTER, J.N. (1972): Ranked set sampling theory with order statititcs background. **Biometrics** 28,545-553.

[9] EICHHORN B. H, and HAYRE LS (1983): Scrambled **randomized response methods for obtaining sensitive question data. Journal** of Statistical Planning and Inference 7, 307–316.

[10] GREENBERG B.G., KUBLER R.R, and D.G. HORVITZ (1971): Applications of RR technique in obtaining quantitative data. **Journal of the American Statistical Association** 66, 243–250.

[11] GUPTA S., SHABIR J., SOUSA R. and CORTE-REAL P. (2012): Estimation of the mean of a sensitive variable in the presence of auxiliary information. **Communications in Statistics-Theory and Methods** 41, 2394–2404.
[12] HUANG K.C. (2010): Unbiased estimators of mean, variance and sensitivity level for quantitative characteristics in finite population sampling. **Metrika** 71, 341–352.
[13] HUSSAIN Z, AL-SOBHI M.M. and AL-ZAHRANI B. (2014): Additive and Subtractive Scrambling in Optional Randomized Response Modeling. **PLoS ONE** 9(1): e83557. doi:10.1371/journal.pone.0083557
[14] MCINTYRE G. A. (1952): A method of unbiased selective sampling using ranked sets. **Australian J. Agricultural Researc**h 3, 385-390.
[15] MEHTA S., DASS B.K., SHABBIR J. and GUPTA S. (2012): A three stage optional randomized response model. **Journal of Statistical Theory and Practice** 6, 417–426.
[16] PATIL G.P., SINHA A.K. and TAILLIE, C.(1995): Ranked set sampling coherent ranking and size biased permutations. **J. Statistical Planning and Inference**.. 63, 311-324.
[17] RUEDA M. and S. GONZÁLEZ (2004): Missing data and auxiliary information in surveys. **Comput. Statist**. 19, 551-567.
[18] RYU, J. B. , J. M. KIM, T. Y. HEO and C. G. PARK (2005): On stratified randomized response sampling, **Model Assist. Stat. Appl**. 1, 31-36.
[19] SOBERANIS CRUZ, V.H. and CUEVAS DOMÍNGUEZ J. D. (2010): Muestreo de respuesta aleatorizadas con probabilidades desiguales: el estimador de Rao-Hartley-Cochran. **Investigación Operacion**al, 31, 3, 248-257.
[20] STOKES, L. (1977): Ranked set sampling with concomitant variables. **Communication in Statistics, Theory and Methods,** 6, 1207- 1211.
[21] TAKAHASHI K. and WAKIMOTO, M. (1968): On unbiased estimates of the population mean based on the sample stratified by means ordering. **Annals of the Institute of Mathematical Statistics**. 20, 1031-1047.
[22] WARNER S.L. (1965): Randomized response: a survey for eliminating evasive answer bias. **Journal of the American Statistical Association** 60, 63–69.