

9-5-2018

Dealing with Sensitive Quantitative Variables: A Comparison of Sampling Designs for the Procedure of Gupta and Thornton


Carlos Narciso Bouza Herrera

University of Havana, bouza@matcom.uh.cu

Prayas Sharma

University of Petroleum and Energy Studies, Dehradun, India

Follow this and additional works at: <https://digitalcommons.wayne.edu/jmasm>

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Bouza Herrera, Carlos Narciso and Sharma, Prayas (2018) "Dealing with Sensitive Quantitative Variables: A Comparison of Sampling Designs for the Procedure of Gupta and Thornton," *Journal of Modern Applied Statistical Methods*: Vol. 17 : Iss. 1 , Article 24.

DOI: 10.22237/jmasm/1536167903

Available at: <https://digitalcommons.wayne.edu/jmasm/vol17/iss1/24>

This Regular Article is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized editor of DigitalCommons@WayneState.

Dealing with Sensitive Quantitative Variables: A Comparison of Sampling Designs for the Procedure of Gupta and Thornton

Carlos Narciso Bouza Herrera

University of Havana
Havana, Cuba

Prayas Sharma

University of Petroleum and Energy Studies
Dehradun, India

The use of randomized response procedures allows diminishing the number of non-responses and increasing the accuracy of the responses. A new sampling strategy is developed where the reports are scrambled using the procedure of Gupta and Thornton. The estimator of the mean as well as the errors are developed for the Rao-Hartley-Cochran and Ranked Sets Sampling designs. The proposals are compared with the original model based on the use of simple random sampling.

Keywords: Randomized responses, Rao-Hartley-Cochran, ranked sets sampling, scrambled variable

Introduction

Interviewers are interested in obtaining honest responses from respondents, but it is difficult when dealing with sensitive issues. Warner's (1965) work on randomized response (RR) models is seminal. It dealt with the estimation of a proportion of positive responses to a sensitive question in a population, and avoided declaring the real status of the respondent. RR models are recommended for both decreasing evasive answer bias and providing privacy protection to the respondents. It is expected that RR increases the response rate and reduce the response error.

Gupta and Thornton (2002) proposed a RR procedure, known as the Gupta-Thornton (GT) RR procedure, which provides confidence to the respondents and possesses interesting properties. It uses two randomization steps. They developed the statistical model using simple random sampling (SRS). Hussain (2012)

extended their procedure introducing the use of two possible questions. The purpose of this study is to extend the model for two sampling designs: Rao-Hartley-Cochran unequal sampling designs and ranked set sampling.

Rao, Hartley, and Cochran (1962) proposed an unequal probability selection procedure. The Rao-Hartley-Cochran (RHC) estimator using RR was studied by Soberanis Cruz and Cuevas Domínguez (2010). McIntyre (1952) proposed a cost-effective survey sampling method that is currently known as ranked set sampling (RSS) in the literature. In this method the sampling units are partitioned into small subsets of the same size. The units of each subset are ranked separately with respect to the characteristic of interest using a concomitant variable. Ranking is supposed to be easily made at a low cost. RSS is an alternative sample design which provides gains in accuracy with respect to simple random sampling with replacement (SRSWR). This theme is of growing importance; recent papers in this context are Al-Omari and Jaber (2008) and Al-Nasser (2007). See Chen, Bai, and Sinha (2004) and Bouza-Herrera (2013) for a detailed discussion on RSS.

Gupta-Thornton RR Procedures

Gupta and Thornton (2002) proposed a RR procedure based on a two-step randomization mechanism. In addition to the sensitive variable Y the surveyor determines a probability density function $f(x)$. A non-sensitive variable X is generated according to $f(x)$. As the sampler fixes $f(x)$, $E(X) = \mu_X \in \mathfrak{R}_y$ and $V(X) = \sigma_X^2 \in \mathfrak{R}^+$ are known. The sampler also fixes a randomizer that generates independent Bernoulli distributed β with $E(\beta) = T$. In the first stage, the interviewer generates a value of X . In the second stage, they generate a value of β . When $\beta = 1$, they report the true value of Y ; in other case the report is $Z = X + Y$. Therefore, the report is the random variable

$$R_i = \beta Y_i + (1 - \beta) Z_i, \quad i = 1, \dots, n$$

Under this RR procedure its expectation is

$$E_M(R_i) = T Y_i + (1 - T)(X_i + Y_i), \quad i = 1, \dots, n$$

Considering the design is SRSWR,

$$E(E_M(R_i)) = T\mu_Y + (1-T)(\mu_X + \mu_Y), \quad i = 1, \dots, n$$

Therefore, the mean of Y is estimated unbiasedly by

$$\hat{\mu}_Y = \bar{R} - (1-T)\mu_Y = \frac{1}{n} \sum_{i=1}^n R_i + (1-T)\mu_Y$$

Its variance is (see Gupta & Thornton, 2002; Gupta, Gupta, & Singh, 2002)

$$V(\hat{\mu}_Y) = \frac{\sigma_Y^2}{n} + \frac{(1-T)(\sigma_X^2 + T\mu_X^2)}{n}$$

Note

$$V(\hat{\mu}_X) = \frac{\sigma_X^2}{n} + \frac{(1-T)(\sigma_Y^2 + T\mu_Y^2)}{n}$$

This model was improved by Hussain (2012). The new model is based on the selection of two responses from each respondent. Each response was used for computing an estimation. They are correlated but have equal variances. The new procedure, identified as Hussain-Gupta-Thornton (H-GT), is described below.

Revised Gupta and Thornton RR (H-GT)

Fix a randomization mechanism (RM) that generates independent Bernoulli variables β with probability T . Fix a mechanism that generates a random variable Y with density $f(y)$. The respondent i is requested to use $f(y)$ and generates two values of Y , Y_{ij} , $j = 1, 2$.

The respondent uses RM for selecting between:

- (i) Report the true response on sensitive variable Y with probability T .
- (ii) Report $Z_{ij} = Y_i + Y_{ij}$ with probability $1 - T$, $j = 1, 2$.

Now, each respondent's reports are modeled by

$$R_{ij} = \beta_j X_i + (1 - \beta_j) Z_{ij}, \quad i = 1, \dots, n; j = 1, 2$$

Consider the case $E(\beta_j) = T, j = 1, 2$. Two estimators are used:

$$\hat{\mu}_{(X+)} = \frac{1}{n} \sum_{i=1}^n R_{i1} + (1-T)\mu_Y, \quad \hat{\mu}_{(X-)} = \frac{1}{n} \sum_{i=1}^n R_{i2} - (1-T)\mu_Y$$

Both are unbiased.

An estimator of the mean of the sensitive variable is obtained by giving weight to each of them. The proposal of Hussein (2012) was

$$\hat{\mu}_{XW} = W\hat{\mu}_{X+} + (1-W)\hat{\mu}_{X-}, \quad W \in]0,1[$$

Its variance is

$$V(\hat{\mu}_{XW}) = W^2V(\hat{\mu}_{X+}) + (1-W)^2V(\hat{\mu}_{X-}) + 2W(1-W)\text{Cov}(\hat{\mu}_{X+}, \hat{\mu}_{X-})$$

It is readily obtained that both variances are equal to

$$V(\hat{\mu}_X) = \frac{\sigma_X^2}{n} + \frac{(1-T)(\sigma_Y^2 + T\mu_Y^2)}{n}$$

However,

$$\begin{aligned} \text{Cov}(\hat{\mu}_{X+}, \hat{\mu}_{X-}) &= \frac{1}{n^2} \sum_{i=1}^n \sum_{i^*=1}^n \text{Cov}(R_{i1}, R_{i^*2}) \\ &= \frac{1}{n^2} \left(\sum_{i \neq i^*}^n \text{Cov}(R_{i1}, R_{i^*2}) + \sum_{i^*=1}^n \text{Cov}(R_{i1}, R_{i^*2}) \right) \\ &= \frac{1}{n^2} \sum_{i^*=1}^n \text{Cov}(R_{i1}, R_{i^*2}) \end{aligned}$$

Note

$$\begin{aligned} \text{Cov}(R_{i1}, R_{i2}) &= E(R_{i1}R_{i2}) - E(R_{i1})E(R_{i2}) \\ &= \sigma_X^2 + (1-T)(\sigma_Y^2 + T\mu_Y^2) \end{aligned}$$

so

$$\text{Cov}(\hat{\mu}_{X+}, \hat{\mu}_{X-}) = \frac{\sigma_X^2}{n} - \frac{(1-T)(\sigma_Y^2 + T\mu_Y^2)}{n}$$

Therefore,

Proposition: Using H-GT for obtaining responses from a SRSWR of n individuals of a finite population (see Hussein, 2012).

- 1) An unbiased estimator of μ_Y is $\hat{\mu}_{XW} = W\hat{\mu}_{X+} + (1-W)\hat{\mu}_{X-}$, $W \in]0,1[$.
- 2) Its variance is $V(\hat{\mu}_X) = \sigma_X^2/n$.

This proposition establishes that H-GT does not increase the sampling error by using its scrambling variable in the randomization response procedure.

RSS Gupta-Thornton RR Procedures

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 individuals in the sample are ranked. Take $Y_{(t:1)}, \dots, Y_{(t:t)}, \dots, Y_{(t:m)}$ as the order statistics (OS) of the sample s_t . Measure only the OS $Y_{(1:1)}, \dots, Y_{(t:t)}, \dots, Y_{(m:m)}$. 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

$$\bar{x}_{\text{rss}} = \frac{1}{mr} \sum_{i=1}^m \sum_{k=1}^r X_{(tr)k}$$

Because $E(X_{(t:t)k}) = \mu_{X(t)}$, $t = 1, \dots, m$, $k = 1, \dots, r$, and

$$\mu_X = \frac{1}{m} \sum_{t=1}^m \mu_{X(t)}$$

it is found that \bar{x}_{rss} estimates μ_Y unbiasedly. Due to independence,

$$V(\bar{x}_{\text{rss}}) = \frac{1}{m^2 r} \sum_{t=1}^m \sigma_{X(t)}^2$$

the relation

$$\sigma_{X(t)}^2 = \sigma_X^2 - \Delta_{X(t)}^2, \quad \Delta_{X(t)} = \mu_{X(t)} - \mu_X$$

permits Y such that

$$V(\bar{x}_{\text{rss}}) = \frac{\sigma_X^2}{mr} - \frac{1}{m^2 r} \sum_{t=1}^m \Delta_{X(t)}^2$$

Assume that there is an expensive method for obtaining information for predicting Y for every sampled person u_i . Rank the selected individual without interviewing them. For example, examining medical records of a selected person permits ranking the possible level of their consumption of drugs Y using a concomitant variable. Stokes (1977) considered the effect of the ranking errors due to the use of Y and determined it does not affect the main statistical properties of the RSS mean estimator.

The report of a sampled individual will be

$$R_{(t)t}k} = \beta X_{(t)t}k} + (1 - \beta) Z_{(t)t}k}, \quad t = 1, \dots, m; k = 1, \dots, r$$

where

$$Z_{(t)t}k} = X_{(t)t}k} + Y_{tk}$$

Y_{tk} is the result of generating a value of Y using $f(y)$ by the respondent ranked t in the RSS sample t in the cycle k .

Then

Proposition: If GT is used for obtaining responses from an RSS selected from a finite population using an auxiliary variable A correlated with Y , considering the $n = mr$ yields

$$R_{(t)t}k} = \beta X_{(t)t}k} + (1 - \beta) Z_{(t)t}k}, \quad t = 1, \dots, m; k = 1, \dots, r$$

1) An unbiased estimator of μ_Y is

DEALING WITH ALTERNATIVES TO GUPTA AND THORNTON

$$\hat{\mu}_{X(\text{rss})} = \bar{R}_{\text{rss}} - (1-T)\mu_Y = \frac{1}{mr} \sum_{t=1}^m \sum_{k=1}^r R_{(t)k} - (1-T)\mu_Y$$

2) Its variance is

$$V(\hat{\mu}_{X(\text{rss})}) = \frac{\sigma_X^2}{n} + \frac{(1-T)(\sigma_Y^2 + T\mu_Y^2)}{n} - \frac{1}{m^2 r} \sum_{t=1}^m \Delta_{X(t)}^2$$

Proof: As

$$\begin{aligned} E(R_{(t)k}) &= T\mu_X + (1-T)(\mu_X + \mu_Y) \\ &= T\mu_X + (1-T)\mu_Y, \quad t = 1, \dots, m; k = 1, \dots, r \end{aligned}$$

$E(\hat{\mu}_{X(\text{rss})}) = \mu_X$ and the unbiasedness of the estimator is derived.

However,

$$V(R_{(t)k}) = E(R_{(t)k}^2) - (T\mu_X + (1-T)\mu_Y)^2$$

The first term is

$$\begin{aligned} E(R_{(t)k}^2) &= E(\beta^2 X_{(t)k} + (1-\beta)^2 Z_{(t)k} + 2\beta(1-\beta) X_{(t)k} Z_{(t)k}) \\ &= T(\mu_X^2 + \sigma_{X(t)}^2) + (1-T)(\mu_X^2 + \sigma_{X(t)}^2 + \mu_Y^2 + \sigma_Y^2 - 2\mu_X \mu_Y) \end{aligned}$$

for $t = 1, \dots, m; k = 1, \dots, r$. Therefore

$$\begin{aligned} V(R_{(t)k}) &= (\mu_X^2 + \sigma_{X(t)}^2) + (1-T)(\mu_X^2 + \sigma_{X(t)}^2 + \mu_Y^2 + \sigma_Y^2 - 2\mu_X \mu_Y) - (T\mu_X + (1-T)\mu_Y)^2 \\ &= \sigma_{X(t)}^2 + (1-T)(\sigma_Y^2 + T\mu_Y^2) \end{aligned}$$

and

$$V\left(\hat{\mu}_{X(\text{rss})}\right) = \frac{\sigma_X^2}{n} + \frac{(1-T)(\sigma_Y^2 + T\mu_Y^2)}{n} - \frac{1}{m^2 r} \sum_{t=1}^m \Delta_{X(t)}^2$$

The last term in $V\left(\hat{\mu}_{X(\text{rss})}\right)$ is the gain in accuracy due to the use of RSS.

In some applications the members of the k^{th} SRSWR selected sample may be convinced to share the values of the generated values of Y . Then the ranking is made on $Y = A$, and work may be done with OSs of Y . If Y is not correlated with A , the unbiasedness of $\hat{\mu}_{X(\text{rss})}$ holds but

$$V\left(\hat{\mu}_{X(\text{rss})} | Y\right) = \frac{\sigma_X^2}{mr} + \frac{(1-T)\left(\sigma_Y^2 - \frac{1}{m^2 r} \sum_{t=1}^m \Delta_{Y(t)}^2 + T\mu_Y^2\right)}{n}, \quad \Delta_{Y(t)} = \mu_{Y(t)} - \mu_Y$$

This result can be considered as a Corollary to the previous propositions.

Corollary: Under the conditions of the above proposition, if $A = Y$ is used for ranking:

1. If Y is uncorrelated with Y

$$R_{(t:t|Y)k} = \beta X_{tk} + (1-\beta)(X_{tk} + Y_{(t:t)k}), \quad t = 1, \dots, m; k = 1, \dots, r$$

- 1.1) An unbiased estimator of μ_Y is

$$\hat{\mu}_{X(\text{rss}|Y)} = \frac{1}{mr} \sum_{t=1}^m \sum_{k=1}^r R_{(t:t|Y)k} - (1-T)\mu_Y$$

- 1.2) Its variance is

$$V\left(\hat{\mu}_{X(\text{rss}|Y)}\right) = \frac{\sigma_X^2}{n} + \frac{(1-T)\left(\sigma_Y^2 + T\mu_Y^2 - \frac{1}{m^2 r} \sum_{t=1}^m \Delta_{Y(t)}^2\right)}{n}$$

2. If Y is correlated with Y

DEALING WITH ALTERNATIVES TO GUPTA AND THORNTON

$$R_{(t)Yk} = \beta^2 X_{(t)k} + (1-\beta) \left(X_{(t)k} + Y_{(t)k} \right), \quad t=1, \dots, m; k=1, \dots, r$$

2.1) An unbiased estimator of μ_Y is

$$\hat{\mu}_{X(\text{rss}Y)} = \frac{1}{mr} \sum_{t=1}^m \sum_{k=1}^r R_{(t)Yk} - (1-T) \mu_Y$$

2.2) Its variance is

$$V\left(\hat{\mu}_{X(\text{rss}Y)}\right) = \frac{\sigma_X^2}{n} + \frac{(1-T)(\sigma_Y^2 + T\mu_Y^2)}{n} - \left(\frac{1}{nr} \sum_{t=1}^m \Delta_{X(t)}^2 + (1-T) \left(\frac{1}{r} \sum_{t=1}^m \Delta_{Y(t)}^2 \right) \right)$$

These results suggest that using a probability function with large values of $|\Delta_{Y(t)}|$ increases the gain in accuracy both in the correlated and uncorrelated cases.

An RSS extension of H-GT is given by

$$R_{(t)k1} = \beta X_{(t)k} + (1-\beta) Z_{(t)k} \quad R_{(t)k2} = \beta X_{(t)k} - (1-\beta) Z_{(t)k}$$

The corresponding unbiased RSS-estimators are

$$\hat{\mu}_{X(\text{rss}+)} = \frac{1}{mr} \sum_{t=1}^m \sum_{k=1}^r R_{(t)k1} - (1-T) \mu_Y$$

$$\hat{\mu}_{X(\text{rss}-)} = \frac{1}{mr} \sum_{t=1}^m \sum_{k=1}^r R_{(t)k2} + (1-T) \mu_Y$$

μ_Y is unbiasedly estimated by

$$\hat{\mu}_{X(\text{rss}W)} = W \hat{\mu}_{X(\text{rss}+)} + (1-W) \hat{\mu}_{X(\text{rss}-)}$$

Two cases are now considered:

Case 1. $A \neq Y$

$$V\left(\hat{\mu}_{X(\text{rSSW})}\right) = \frac{\sigma_X^2}{n} - \frac{1}{nm} \sum_{t=1}^m \Delta_{X(t)}^2$$

Case 2. $A = Y$

$$V\left(\hat{\mu}_{X(\text{rSSW}|Y)}\right) = \frac{\sigma_X^2}{n}$$

because the ranking of Y is random.

RHC Sample Selection for Applying Gupta-Thornton RR Procedures

Rao et al. (1962) proposed using a set of unequal probabilities for selecting a sample of size n . An auxiliary variable X is used for determining the set of selection probabilities. Their drawing procedure is the following:

Drawing Procedure of Rao-Hartley-Cochran

Divide the population $U = \{1, \dots, N\}$ into n disjoint sets $U_i, i = 1, 2, \dots, n$, of size N_i each. Select a unit k_i from U_i with probability

$$\frac{x_{k_i}}{t_{x_i}}, \quad t_{x_i} = \sum_{t=1}^{N_i} x_{it}$$

Repeat for $i < n$.

Consider the Rao-Hartley-Cochran (RHC) model when using the GT procedure. Calculate the corresponding conditional expectation under each procedure and deal with the basic formulae to derive adequate estimators.

Substituting the scrambled variable in the usual RHC estimator, we have that the total, when the GT procedure is used, is given by:

$$\hat{t}_{AI, \text{RHC}} = \sum_{i=1}^n \frac{B_{k_i}}{p_{k_i}} \pi_i$$

The conditional expectation of this estimator is

DEALING WITH ALTERNATIVES TO GUPTA AND THORNTON

$$\begin{aligned}
 E_{RC}(\hat{t}_{AI,RHC}) &= E_{RC}\left(\sum_{i=1}^n B_i \frac{\pi_i}{p_i}\right) \\
 &= \sum_{i=1}^n E_{RC}(B_i) \frac{\pi_i}{p_i} \\
 &= \sum_{i=1}^n E_{RC}(TY_i + (1-T)(X_i + Y_i)) \frac{\pi_i}{p_i} \\
 &= \sum_{i=1}^n \mu_Y + (1-T)\mu_X \frac{\pi_i}{p_i}
 \end{aligned}$$

Hence

$$\begin{aligned}
 E(\hat{t}_{AI,RHC}) &= E_1 \left\{ E_2 \left[E_{RC}(\hat{t}_{AI,RHC}) \right] \right\} \\
 &= E_1 E_2 \left(\sum_{i=1}^n \mu_Y + (1-T)\mu_X \frac{\pi_i}{p_i} \right) \\
 &= N(\mu_Y + (1-T)\mu_X)
 \end{aligned}$$

Therefore, an estimator of the population mean is

$$\bar{y}_{RHC} = \frac{1}{N} \sum_{i=1}^n \frac{B_{k_i}}{p_{k_i}} \pi_i - (1-T)\mu_X$$

It is unbiased for the estimation of μ_Y . Take

$$\begin{aligned}
 N^2 \cdot V(\hat{t}_{AI,RHC}) &= E_1 \left\{ E_2 \left[V(\hat{t}_{AI,RHC}) \right] \right\} + E_1 \left\{ V_2 \left[E_{RC}(\hat{t}_{AI,RHC}) \right] \right\} + V_1 \left\{ E_2 \left[E_{RC}(\hat{t}_{AI,RHC}) \right] \right\} \\
 &= E_1 \left\{ V_2 \left[E_{RC}(\hat{t}_{AI,RHC}) \right] \right\} + V_1 \left\{ E_2 \left[E_{RC}(\hat{t}_{AI,RHC}) \right] \right\} + E_1 \left\{ E_2 \left[V_{RC}(\hat{t}_{AI,RHC}) \right] \right\} \\
 &= E_1 V_2 \left(\sum_{i=1}^n \mu_Y + (1-T)\mu_X \frac{\pi_i}{p_i} \right) + E_1 \left\{ E_2 \left[V_{RC}(\hat{t}_{AI,RHC}) \right] \right\}
 \end{aligned}$$

and introduce the Bernoulli variables

$$I_{G_{i,t}} = \begin{cases} 1 & \text{if } t \in G_i; \\ 0 & \text{otherwise} \end{cases}, \quad I_{G_{i,t,t'}} = \begin{cases} 1 & \text{if } t, t' \in G_i; \\ 0 & \text{otherwise} \end{cases}$$

Their conditional expectations are

$$E_1(I_{G_{i,t}}) = \Pr\{I_{G_{i,t}} = 1\} = \frac{N_i}{N}$$

and

$$E_1(I_{G_{i,t,t'}}) = \Pr\{I_{G_{i,t,t'}} = 1\} = \Pr\{t \& t' \in G_i\} = \frac{N_i(N_i - 1)}{N(N - 1)}$$

Now derive

$$E_1 V_2 \left(\sum_{i=1}^n \mu_Y + (1-T)\mu_X \frac{\pi_i}{p_i} \right) = (\mu_Y + (1-T)\mu_X)^2 \frac{\sum_{i=1}^n N_i^2 - N}{N(N-1)}$$

Note

$$V_{RC}(\hat{t}_{A1,RHC}) = \sum_{i=1}^n V_{RC}(B_i) \left(\frac{\pi_i}{p_i} \right)^2 = \sum_{i=1}^n V_{0i} \left(\frac{\pi_i}{p_i} \right)^2$$

where $V_{0i} = V_0 = \sigma_Y^2 + (1-T)(\sigma_X^2 + T\mu_X^2)$. Using the results of Soberanis Cruz and Cuevas Domínguez (2010), derive

$$\begin{aligned} E_2[V_{RC}(\hat{t}_{A1,RHC})] &= V_0 \sum_{i=1}^n E_2 \left[\left(\frac{\pi_i}{p_i} \right)^2 \right] \\ &= V_0 \sum_{i=1}^n \left\{ \sum_{t=1}^{N_i} \frac{\pi_i^2 p_t}{p_t^2 \pi_i} \right\} \\ &= V_0 \sum_{i=1}^n \left\{ \sum_{t=1}^{N_i} \frac{\pi_i}{p_t} \right\} \end{aligned}$$

DEALING WITH ALTERNATIVES TO GUPTA AND THORNTON

$$\begin{aligned}
 E_2 \left[V_{RC} \left(\hat{t}_{AI,RHC} \right) \right] &= V_0 \sum_{i=1}^n \pi_i \sum_{t=1}^{N_i} \frac{1}{p_t} \\
 &= V_0 \sum_{i=1}^n \left(\sum_{t=1}^N p_t I_{G_i;t} \right) \left(\sum_{t'=1}^N \frac{1}{p_{t'}} I_{G_i;t'} \right) \\
 &= V_0 \sum_{i=1}^n \left(\sum_{t=1}^N I_{G_i;t}^2 + \sum_{t=1}^N \sum_{\substack{t'=1 \\ t' \neq t}}^N \frac{p_t}{p_{t'}} I_{G_i;t,t'} \right) \\
 &= V_0 \sum_{i=1}^n \left(N_i + \sum_{t=1}^N \sum_{\substack{t'=1 \\ t' \neq t}}^N \frac{p_t}{p_{t'}} I_{G_i;t,t'} \right) \\
 &= V_0 \left[N + \sum_{i=1}^n \left(\sum_{t=1}^N \sum_{\substack{t'=1 \\ t' \neq t}}^N \frac{p_t}{p_{t'}} I_{G_i;t,t'} \right) \right]
 \end{aligned}$$

Therefore

$$\begin{aligned}
 E_1 \left\{ E_2 \left[V_{RC} \left(\hat{t}_{AI,RHC} \right) \right] \right\} &= V_0 \left[N + \sum_{i=1}^n E_1 \left(\sum_{t=1}^N \sum_{\substack{t'=1 \\ t' \neq t}}^N \frac{p_t}{p_{t'}} I_{G_i;t,t'} \right) \right] \\
 &= V_0 \left[N + \sum_{i=1}^n \left(\frac{N_i(N_i-1)}{N(N-1)} \sum_{t=1}^N \sum_{\substack{t'=1 \\ t' \neq t}}^N \frac{p_t}{p_{t'}} \right) \right] \\
 &= V_0 \left[N + \frac{\sum_{i=1}^n N_i^2 - N}{N(N-1)} \sum_{t=1}^N \sum_{\substack{t'=1 \\ t' \neq t}}^N \frac{p_t}{p_{t'}} \right]
 \end{aligned}$$

As

$$\begin{aligned}
 &V \left(\hat{t}_{AI,RHC} \right) \\
 &= N^{-2} \left[\frac{\sum_{i=1}^n N_i^2 - N}{N(N-1)} \right] (\mu_Y + (1-T)\mu_X)^2 + V_0 \left[N + \frac{\sum_{i=1}^n N_i^2 - N}{N(N-1)} \sum_{t=1}^N \sum_{\substack{t'=1 \\ t' \neq t}}^N \frac{p_t}{p_{t'}} \right]
 \end{aligned}$$

Dividing this expression by $(N(P + Q\mu_z))^2$,

$$V(\bar{y}_{\text{RHC}}) = \frac{\sum_{i=1}^n N_i^2 - N}{N^3(N-1)} + \frac{V_0 \left[N + \frac{\sum_{i=1}^n N_i^2 - N}{N(N-1)} \sum_{t=1}^N \sum_{\substack{t'=1 \\ t' \neq t}}^N p_t p_{t'} \right]}{N^2}$$

It is acceptable even for moderate values of N

$$F = \frac{\sum_{i=1}^n N_i^2 - N}{N^3(N-1)} \cong \frac{\sum_{i=1}^n N_i^2}{N^4} - \frac{1}{N^3} \rightarrow 0$$

Then, in many real life problems,

$$V(\bar{y}_{\text{RHC}}) \rightarrow \frac{V_0}{N}$$

When this approximation is valid, it is preferred to use RHC instead of SRSWR, for the GT-RR method, because

$$\frac{V(\hat{\mu}_y)}{V(\bar{y}_{\text{RHC}})} \cong \frac{N}{n} \geq 1$$

Conclusions

An alternative sample strategy was developed whereby using ranked set sampling (RSS) is a better alternative than simple random sampling with replacement in terms of the accuracy of the estimators. The Rao-Hartley-Cochrane (RHC) sampling design may also increase the accuracy. Under some conditions it is expected that RHC outperforms the gain in accuracy of RSS. This seems to be a consequence of the stratification provoked by the use of both procedures. RSS is to be preferred to RHC only if there is additional information that allows obtaining a non-random ranking of the sensitive variable.

Acknowledgements

One of the authors thanks for the support of Red CYTED 516RT0513.

References

- Al-Nasser, D. A. (2007). L ranked set sampling: A generalization procedure for robust visual sampling. *Communication in Statistics – Simulation and Computation*, 36(1), 33-44. doi: 10.1080/03610910601096510
- Al-Omari, A. I., & Jaber, K. H. (2008). Percentile double ranked set sampling. *Journal of Mathematics and Statistics*, 4(1), 60-64. doi: 10.3844/jmssp.2008.60.64
- Bouza-Herrera, C. N. (2013). *Handling with missing data in ranked set sampling*. New York, NY: Springer.
- Chen, Z., Bai, Z., & Sinha, B. K. (2004). *Ranked set sampling: Theory and applications*. New York, NY: Springer. doi: 10.1007/978-0-387-21664-5
- Gupta, S., Gupta, B., & Singh, S. (2002). Estimation of sensitivity level of personal interview survey questions. *Journal of Statistical Planning and Inference*, 100(2), 239-247. doi: 10.1016/s0378-3758(01)00137-9
- Gupta, S., & Thornton, B. (2002). Circumventing social desirability response bias in personal interview surveys. *American Journal of Mathematical and Management Sciences*, 22(3-4), 369-383. doi: 10.1080/01966324.2002.10737599
- Hussain, Z. (2012). Improvement of the Gupta and Thornton scrambling model through double use of randomization device. *International Journal of Academic Research in Business and Social Sciences*, 2(6), 91-97.
- McIntyre, G. A (1952). A method for unbiased selective sampling, using ranked sets. *Australian Journal of Agricultural Research*, 3(4), 385-390. doi: 10.1071/ar9520385
- Rao, J. N. K., Hartley, H. O., & Cochran, W. G. (1962). On a simple procedure of unequal probability sampling without replacement. *Journal of the Royal Statistical Society. Series B (Methodological)*, 24(2), 482-491. Available from <https://www.jstor.org/stable/2984239>
- Soberanis Cruz, V. H., & Cuevas Domínguez, J. D. (2010). Muestreo de respuesta aleatorizadas con probabilidades desiguales: El estimador de Rao-Hartley-Cochran [Randomized response sampling with unequal probabilities: The

Rao-Hartley-Cochran estimator]. *Investigación Operacional*, 31(3), 248-257.
Retrieved from

<http://ojs.uh.cu/InvestigacionOperacional/index.php/InvOp/article/view/361>

Stokes, S. L. (1977). Ranked-set sampling with concomitant variables. *Communications in Statistics – Theory and Methods*, 6(12), 1207-1212. doi: 10.1080/03610927708827563

Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309), 63-69. doi: 10.1080/01621459.1965.10480775